Star Clusters Across Cosmic Time

Mark R. Krumholz[1,2], Christopher F. McKee[3], and Joss Bland-Hawthorn[2,4,5]

[1]Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia; email: mark.krumholz@anu.edu.au
[2]Centre of Excellence for All Sky Astrophysics in Three Dimensions (ASTRO-3D), Australia
[3]Departments of Astronomy and of Physics, University of California, Berkeley, CA 94720, USA
[4]Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW 2006, Australia
[5]Miller Professor, Miller Institute, University of California, Berkeley, CA 94720, USA

1

Keywords

Abstract

Star clusters stand at the intersection of much of modern astrophysics:
the interstellar medium, gravitational dynamics, stellar evolution, and
cosmology. Here we review observations and theoretical models for the
formation, evolution, and eventual disruption of star clusters. Current
literature suggests a picture of this life cycle with several phases:

- Clusters form in hierarchically-structured, accreting molecular
  clouds that convert gas into stars at a low rate per dynamical
  time until feedback disperses the gas.
- The densest parts of the hierarchy resist gas removal long enough
  to reach high star formation efficiency, becoming dynamically-
  relaxed and well-mixed. These remain bound after gas removal.
- In the first ~ 100 Myr after gas removal, clusters disperse mod-
  erately fast, through a combination of mass loss and tidal shocks
  by dense molecular structures in the star-forming environment.
- After ~ 100 Myr, clusters lose mass via two-body relaxation and
  shocks by giant molecular clouds, processes that preferentially
  affect low-mass clusters and cause a turnover in the cluster mass
  function to appear on $\sim 1 - 10$ Gyr timescales.
- Even after dispersal, some clusters remain coherent and thus de-
  tectable in chemical or action space for multiple galactic orbits.

In the next decade a new generation of space- and AO-assisted
ground-based telescopes will enable us to test and refine this picture.

Contents

1. INTRODUCTION

The galaxy is in fact nothing but a congeries of innumerable stars grouped together in clusters. . .

Siderius Nuncius (Sidereal Messenger), Galileo Galilei, 1610

1.1. Historical background and motivation

After four centuries since Galileo's first observations, star clusters remain mysterious. What are the unifying principles that govern the formation of globular and open clusters, groups, associations and super-associations? As illustrated in Figure 1[1], clusters cover a huge range of mass, size, and density scales. The clusters shown span ages from ~ 1 Myr to > 10 Gyr, and masses from $\sim 10^2 - 10^6$ $M_\odot$. Some are so

---

[1]The source code and data tabulations used to produce all the figures in this review are available at https://bitbucket.org/krumholz/cluster_review/, and also are provided as Supplementary Materials. We make use of the following software packages: scipy (Oliphant 2007; Millman & Aivazis 2011), matplotlib (Hunter 2007), astropy (Astropy Collaboration et al. 2013), and SLUG (da Silva, Fumagalli & Krumholz 2012; Krumholz et al. 2015b).

ONC, $T < 3$ Myr	Arches, $T \approx 2 - 3$ Myr	NGC 265, $T \approx 250$ Myr	Hyades, $T \approx 625$ Myr

Coll 261, $T \approx 7$ Gyr	NGC 6535, $T > 10$ Gyr	47 Tuc, $T > 10$ Gyr	NGC 1252: Fake News
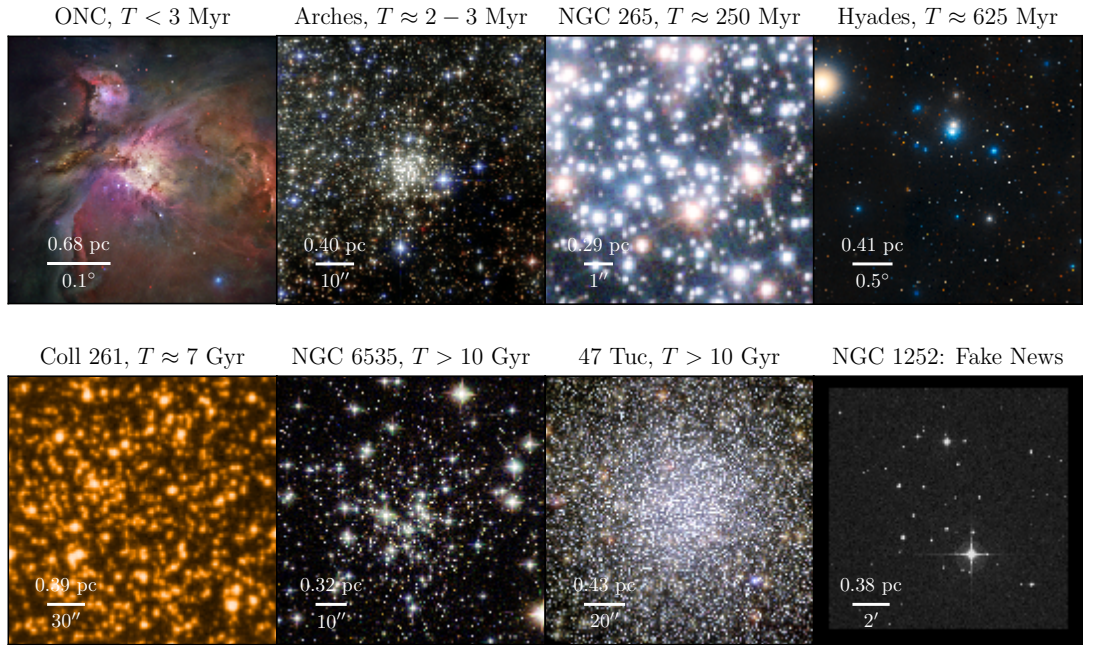
Figure 1

Images of a range of star clusters, along with NGC 1252, an object previously classified as a cluster but now known to be an asterism. The field of view in all frames is 3 pc × 3 pc, and North is up; angular sizes are indicated by scale bars. Image sources are: Orion Nebular Cluster (ONC) – Robberto et al. (2013); Arches cluster – NASA & ESA; NGC 265 – NASA & ESA; Davide De Martin (ESA/Hubble) and Edward W. Olszewski (University of Arizona, USA); Hyades – NASA, ESA, & STScI; Collinder 261 – ESO/Digitized Sky Survey; NGC 6535 – ESA/Hubble & NASA, Gilles Chapdelaine; 47 Tuc – NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration; J. Mack (STScI) and G. Piotto (University of Padova, Italy); NGC 1252 – WEBDA database, https://www.univie.ac.at/webda/.

compact and rich that stars become lost in confusion in the 3 pc-frames shown, others are so sparse and extended that most cluster members are outside the frame. Some are classified as open clusters (Arches, NGC 265, Hyades, Coll 261), some as globular clusters (NGC 6535, 47 Tuc). The Orion Nebula Cluster (ONC) is still forming and, depending on the author, might not even be classified as a cluster at all. NGC 1252 had been classified as an open cluster since 1888, but in 2018 was shown to be merely an asterism (Kos et al. 2018b). Cluster formation is central to the star formation process. Conceivably, *all* stars formed in groups, clusters, or hierarchies, although, for this to be true, most clusters must have dissolved into the Galactic background soon after formation. However, our understanding of when, how, and why stars cluster remains primitive (see reviews by Krumholz et al. 2014, Renaud 2018, and Adamo & Bastian 2018).

A review of clusters is timely for several reasons. One is the emerging overlap between the traditionally separate communities focused on globular clusters and modern-day star clusters and star formation. These have been separate in the past because star clusters in the disk of the Milky Way, generally classified as open clusters (OCs), and those in the halo, generally classified as globular clusters (GCs), appear to occupy distinct loci in the parameter space of mass, age, and metallicity. New data have begun to blur this distinction: even in the Milky Way there is substantial overlap in both metallicity and density between OCs

and GCs, and in rapidly star-forming local galaxies such as the Antennae (Zhang & Fall 1999) and M82 (McCrady & Graham 2007) there is overlap in mass as well. While there is still a bimodality in the star cluster age and spatial distribution, this may simply reflect the combined effects of cluster dissolution and the history of galaxy assembly. In this view, toward which we shall mostly tend here (see also Kruijssen 2014 and Forbes et al. 2018), GCs are no different than any other type of star cluster in their formation and internal dynamics; only their cosmological history is different. Their origins are therefore best addressed in the context of general models for the formation and evolution of star clusters, hence the need for this review.[2]

An additional motivation for this review is to bring together the results of modern studies of stellar kinematics and abundances with the formation and evolution of star clusters. As with GCs and OCs, these have traditionally been somewhat separate fields, with the stellar kinematics and abundances community focused more on the long-term dynamical evolution of stars in and around the Galaxy, and the star formation community paying more attention to the dynamics of interstellar gas and the effects of star formation and stellar feedback on it. This separation is no longer viable. In the era of *Gaia* and massive spectroscopic surveys, kinematic and abundance data are now becoming accurate enough that it should be possible to trace stars that are now part of the field back to their birth places, or at least to reconstruct *some* of the now-dissolved structures in which they were born. To guide such reconstructions, however, will require theoretical and observational input from the star formation community.

With these motivations in mind, we first review current observational constraints on star cluster populations, and then discuss our current understanding, or lack therefore, of the processes by which clusters form, evolve, and eventually disperse.

1.2. Prelude: what is a star cluster?

1.2.1. The need for a definition. The first question that any discussion of star clusters must face, going back at least to Trumpler (1930), is how to define a cluster and distinguish clusters from multiple systems. In their recent review, Portegies Zwart, McMillan & Gieles (2010) define a star cluster as a set of stars that are gravitationally bound to one another, while the earlier review by Lada & Lada (2003) defines a cluster as a collection of stars with a mass density large enough ($\gtrsim 1\ M_\odot\ \mathrm{pc}^{-3}$) to resist tidal disruption in Solar neighborhood conditions, and numerous enough to avoid N-body evaporation for at least 100 Myr. One might also consider defining clusters as concentrations of stars whose mass density significantly exceeds the mean in their galactic neighborhood ($\approx 0.1\ M_\odot\ \mathrm{pc}^{-3}$ near the Sun – McKee, Parravano & Hollenbach 2015). For older stellar populations these definitions are in practice nearly identical. A group of stars of mean density ρ_* and radius r is unbound only if its velocity dispersion $\sigma \gtrsim r\sqrt{G\rho_*}$, so the time required for it to disperse is of order the crossing time $t_{\mathrm{cr}} = r/\sigma \lesssim 1/\sqrt{G\rho_*}$. At the density threshold proposed by Lada & Lada (2003), this is $\lesssim 10$ Myr, so overdensities older than this must be either held together by self-gravity or external forces, or be short-lived transients.

In this review, however, we are interested in the formation of star clusters as well as their evolution, and thus we cannot limit ourselves to stellar populations that are older than their internal dynamical time (c.f. Portegies Zwart, McMillan & Gieles 2010; Gieles & Portegies Zwart 2011). However, younger populations often have complex internal structures such that there is no single way to decompose them into clusters. We illustrate this point with the example of the ONC in Figure 2, where the left two panels show the results of applying two different cluster identification methods that show little agreement: a Bayesian

---

[2]We do note, however that nuclear star clusters probably do represent a physically distinct class, in that their formation and evolution are inseparably linked with those of the central black hole around which they orbit; for this reason we will not include nuclear clusters within the scope of this review.
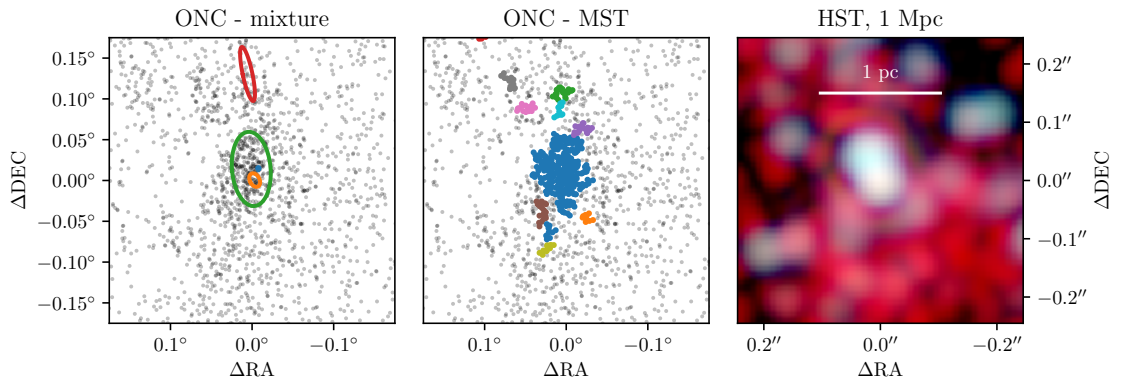
Sample decompositions of the stars in and around the Orion Nebula Cluster (ONC). In the left two panels, gray dots show stars from the catalog of Da Rio et al. (2009); the axes indicate offset from the position of θ^1c Ori, the most luminous star in the cluster. In the left panel, ellipses mark the clusters identified by Kuhn et al. (2014) based on a Bayesian decomposition into isothermal ellipsoids. The middle panel shows a minimum spanning tree (MST) decomposition, performed on the Da Rio et al. (2009) catalog using a pruning length of 0.′3; different colours indicate the different sub-clusters identified by MST. The right panel is a synthetic UBV image produced by placing the ONC to a distance of 1 Mpc (assuming that the true distance is 389 pc – Kounkel et al. 2018), and assuming diffraction-limited imaging with a 2.4 m-diameter telescope such as the *Hubble Space Telescope*. The region shown in the right panel is the same as shown in the two other panels. Colors are scaled logarithmically to give a dynamic range of 100 in flux for all channels.

decomposition into isothermal ellipsoids (Kuhn et al. 2014, left), and a minimum spanning tree (middle). Numerous other decompositions are possible as well (Schmeja 2011). It is not clear which of these, if any, correspond to physically meaningful concepts such as a distinction between structures with positive or negative total energy. Moreover, none of the clustering methods used on resolved stars in the Milky Way are likely to correspond to the clusters picked out by most extragalactic surveys. As the right panel of Figure 2 shows, the ONC would certainly be seen as a single cluster (were it detected at all) in a galaxy at a distance of ≈ 1 Mpc, as targeted by the PHAT survey (Dalcanton et al. 2012, see Margin Note), or $\sim 3 - 10$ Mpc, typical of the LEGUS survey (Calzetti et al. 2015). We return to the question of how to handle the ambiguity in defining clusters at young ages in § 2.7.

1.2.2. Working definition. Given the discussion above, we shall follow Trumpler (1930) and define clusters very generally, so that they can be studied throughout their lives. For our purposes a star cluster is a group of at least 12 stars (to clearly distinguish it from a multiple star system – Trumpler 1930) with a mean density that (i) is at least a factor of a few times the background density (similar to the constraint adopted by Lada & Lada 2003), with the factor being larger in larger galaxies so that the overdensity is statistically significant, and (ii) is much greater than the local density of dark matter. Such groups of stars are physically associated while they form, but there is no constraint on the formation timescale or lifetime; in contrast to Lada & Lada (2003), we do not require the cluster to be denser than the gas cloud out of which it forms, nor do we require that it have enough members to avoid evaporation for 100 Myr. Clusters generally dissipate in time (§ 4), and once they no longer satisfy our criteria we refer to them as "dissolved clusters" (§ 5).

Star clusters can be gravitationally bound or unbound, and young clusters can be a mixture. Unbound

clusters are termed associations; bound clusters can be categorized as OCs or GCs. The most massive and young OCs are sometimes referred to as young massive clusters (YMCs) or super-star clusters (SSCs); they are sometimes also called young GCs, although it is unknown if they will in fact evolve into clusters like the GCs we observe today. YMCs have traditionally been defined by mass $> 10^4\,M_\odot$ and age < 100 Myr, but the lines are fuzzy and there is no particular physical significance to them. Similarly, OCs and GCs do not have precise definitions. In the Milky Way, most OCs have masses $\lesssim 5000\,M_\odot$ and ages $\lesssim 6$ Gyr, whereas most GCs have masses $\gtrsim 10^4\,M_\odot$ and ages $\gtrsim 6$ Gyr (Kharchenko et al. 2013), but in other galaxies these catagories have more overlap. The metallicity of OCs overlaps that of the thick-disk population of Galactic GCs, which has [Fe/H] > -0.8 (Zinn 1985). Clusters with ages $\lesssim$ few hundred Myr are often organized hierarchically, with OCs and YMCs being embedded in associations with sizes up to a few hundred pc, associations grouped together in complexes at $\sim$kpc scales, etc.; there is no physical division between these scales for the unbound structures (Gouliermis 2018). Despite the wide range of physical properties of these various types of star clusters, there is no evidence that they do *not* form via the same basic mechanism (Elmegreen & Efremov 1997; § 3). The only clear intrinsic difference among these different types of clusters is that old GCs exhibit anti-correlations in some light element abundances and have multiple stellar populations, hallmarks that are *never* observed in OCs (§ 2.6). Since these phenomena, which are not understood, have been reviewed recently by Renzini (2013) and Bastian & Lardo (2018), we shall not discuss them further here.

2. DEMOGRAPHICS: OBSERVATIONAL CONSTRAINTS

The physical processes that govern the formation and evolution of star clusters encode themselves in the distributions of star clusters' properties – mass, age, size, etc. Any review of star clusters must therefore begin with the observational constraints and how they are obtained. In this section and the subsequent ones we will define a number of symbols, which we summarize for convenience in Table 1.

2.1. Methods

Methods for determining the demographics of star clusters can be divided into those that operate on resolved stellar populations, and those that operate on unresolved populations.

2.1.1. Resolved Stellar Populations. For star clusters in the Milky Way, the Magellanic Clouds, and M31 (using *Hubble Space Telescope* resolution), it is possible to resolve individual main sequence stars down to masses of a few $M_\odot$ or less. This statement oversimplifies the situation somewhat; the luminosity or mass limit down to which it is possible to resolve stars is a complex function of the stellar density and luminosity distribution in the region being studied. It is possible to resolve massive main sequence stars in relatively sparse regions even beyond the Local Group (e.g., Larsen et al. 2011), while confusion is a problem for the most crowded regions even in the Milky Way (e.g., Ascenso, Alves & Lago 2009). There exists a broad range of parameter space where stars are partially resolved, i.e., their separations are comparable to the observational point spread function (psf); the statistical and photometric techniques used to analyze such fields are beyond the scope of this review.

When individual stars can be resolved, placing them on a color-magnitude diagram (CMD) provides the most direct method of determining the properties of the parent star cluster. In principle, the CMD allows one to read off the masses of the individual stars almost directly and, if the main sequence turn-off is resolved, to infer the age of the population as well. By adding spectroscopy to photometry, one can also determine the stars' compositions. While this method is direct, it does encounter complications. In young clusters where the resolved stars are pre-main sequence, the choice of evolutionary tracks can induce

PDF: probability distribution function

Initial mass function (IMF): PDF of initial mass distribution of stars in a stellar ensemble

Present day mass function (PDMF): PDF of observed mass distribution of stars in a stellar ensemble

Initial cluster mass function (ICMF): PDF of initial mass distribution of star clusters

Cluster mass function (CMF): PDF of observed mass distribution of star clusters

**Cluster age function (CAF):** PDF of observed age distribution of star clusters

Table 1   Symbols used in this review

Symbol	Meaning
α_M	Index of cluster mass distribution, $dN/dM \propto M^{\alpha_M}$ (§ 2.2)
M_c	Upper truncation in cluster mass function (§ 2.2)
α_T	Index of cluster age distribution, $dN/dT \propto T^{\alpha_T}$ (§ 2.3)
Γ	Fraction of stellar mass in bound clusters (§ 2.4)
r_h	Cluster half-mass radius (§ 2.5)
α_{vir}	Virial ratio of a cloud or cluster (§ 3.1)
M_J	Jeans mass (§ 3.1)
t_{cr}	Crossing time of a star cluster (§ 3.1)
t_{ff}	Free-fall time (§ 3.2)
ϵ_{ff}	Fraction of gas converted to stars per free-fall time (§ 3.2)
η	Mass loading factor, i.e., ratio of mass ejected to mass converted to stars (§ 3.2, § 3.3)
ϵ_*	Fraction of gas converted to stars at the end of cluster formation (§ 3.2, § 3.3)
t_{sf}	Time over which star formation takes place (§ 3.2, § 3.4)
v_{esc}	Escape speed from a star cluster (§ 3.3.1)
Σ_{DR}	Surface density below which direct radiation pressure becomes important (§ 3.3.3)
Σ_{IR}	Surface density above which indirect radiation pressure becomes important (§ 3.3.4)
ρ_{SN}	Density above which cluster formation is complete before the first supernova (§ 3.3.6)
σ	Cluster velocity dispersion (§ 4.1)
r_c	King (1966) or Elson, Fall & Freeman (1987) model cluster core radius (§ 4.1)
r_{tr}	King (1966) model cluster truncation radius (§ 4.1)
W_0	King (1966) model cluster dimensionless central potential (§ 4.1)
c	King (1966) model cluster concentration parameter, $c = \log(r_{\mathrm{tr}}/r_c)$ (§ 4.1)
r_{ti}	Cluster tidal radius (§ 4.1)
E_J	Jacobi energy (§ 4.1)
t_{rlx}	Cluster relaxation time at the half-mass radius (§ 4.2.2)

significant uncertainties in the final properties (e.g., Da Rio et al. 2012), and a non-negligible range of ages may be present, so that the system cannot be described by a single age (e.g., Reggiani et al. 2011; Getman et al. 2018). If the observations do not resolve down to the peak of the initial mass function (IMF), one may need to extrapolate using an assumed IMF to account for the mass of unresolved stars (e.g., Lada & Lada 2003; Beerman et al. 2012).

Despite these limitations, however, two other uncertainties are more important. One is the error introduced by the need to correct demographics for clusters that are too small or too old to make it into the catalog. For resolved observations, catalog completeness is invariably a function not just of the mass and age of the stellar population, but also its concentration, the exact masses and ages of its most luminous few stars, and the density of the background (Silva-Villa & Larsen 2011; Johnson et al. 2015b). The other source of uncertainty is in assignment of stars to clusters. As already discussed in § 1.2, for populations that are not old enough to have relaxed, there may be no unique way to decompose a collection of stars into clusters, and variations in how one defines a cluster can lead to order of magnitude differences in the inferred number of clusters and their properties (Bressert et al. 2010). However, even for somewhat older populations, catalogs constructed using different methods can easily differ by $\sim 10 - 20\%$ even when the underlying data are identical (Johnson et al. 2015b). These errors are generally larger than the uncertainties on the properties of individual clusters.
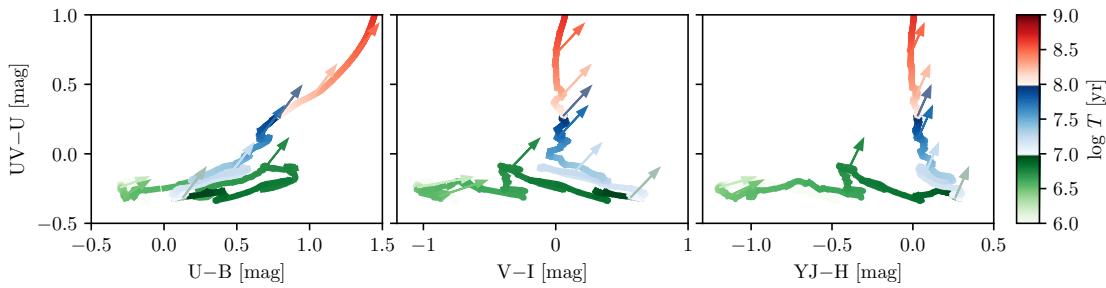
Color evolution of a simple stellar population with a fully sampled IMF, including nebular emission. In each panel the line shows the intrinsic color (in AB mag) as a function of population age. Arrows, placed at intervals of 0.25 dex in age starting at $\log(T/\text{yr}) = 6.25$, show observed color for a visual extinction $A_V = 0.5$ mag for the stars, $A_V = 1.05$ mag for nebular light. This calculation uses SLUG stellar population synthesis code (da Silva, Fumagalli & Krumholz 2012; Krumholz et al. 2015b), run with a Chabrier (2005) IMF, Solar metallicity, MIST rotating stellar evolution models (Choi et al. 2016), starburst99 atmosphere models (Leitherer et al. 1999), a Calzetti et al. (2000) starburst dust attenuation curve, and HST/WFC3 filters. However, other choices for these parameters yield qualitatively similar results.

2.1.2. Unresolved Populations. While CMDs offer the most reliable method for determining star cluster properties, at distances $\gtrsim 1$ Mpc it is not possible to resolve any but the most luminous individual stars. Consequently, studies that probe a wide range of environments must use unresolved populations. In an unresolved observation, a star cluster is distinguished from a single star or other point source either by the fact that its light is more extended than the observational psf (for targets at distances below tens of Mpc) or that it is too luminous to be a single massive star (for more distant targets). In the earliest work with unresolved populations authors simply analyzed luminosity functions (e.g., Whitmore et al. 1993, 1999), but one can also infer physical properties using simple stellar population (SSP) models. In practice this procedure can be carried out in several different ways. The simplest is to generate theoretical luminosities as a function of mass and age by assuming a stellar population that fully samples the IMF (including models for nebular emission and dust extinction) and find the mass and age that best match the observations using χ^2 or a similar goodness-of-fit statistic (e.g., Zhang & Fall 1999; Adamo et al. 2010). More complex methods take into account the stochastic fluctuations in light output that occur as a result of partial sampling of the IMF in clusters smaller than $\sim 10^4 \; M_\odot$ (e.g., Fouesneau & Lançon 2010; Popescu, Hanson & Elmegreen 2012; de Meulenaer et al. 2013; Krumholz et al. 2015a) or add narrowband photometry targeting nebular lines (usually Hα) to provide better age discrimination (e.g., Fouesneau et al. 2012; Bastian et al. 2012; Chandar et al. 2016).

Fitting star cluster properties from photometry leads to substantially larger errors than doing so from resolved measurements of individual stars, with the magnitude of the error depending on the locus in color space. We illustrate this point in Figure 3, which shows the trajectory that an SSP traces out in color over its lifetime. The process of inferring cluster age and extinction essentially amounts to placing an observed point on a multidimensional color-color plot such as Figure 3, and finding the location on an extinction-adjusted SSP track such as the one shown that gets as close as possible to the observed point in every dimension. As is apparent from the figure, for ages below ≈ 5 Myr or above ≈ 50 Myr, in at least some color combinations, SSP tracks are relatively straight and not parallel to the extinction vector. Consequently, the matching process is straightforward and there is little ambiguity in the determination of ages; comparing photometric ages to those inferred from CMDs, Elson & Fall (1985, 1988) estimate that
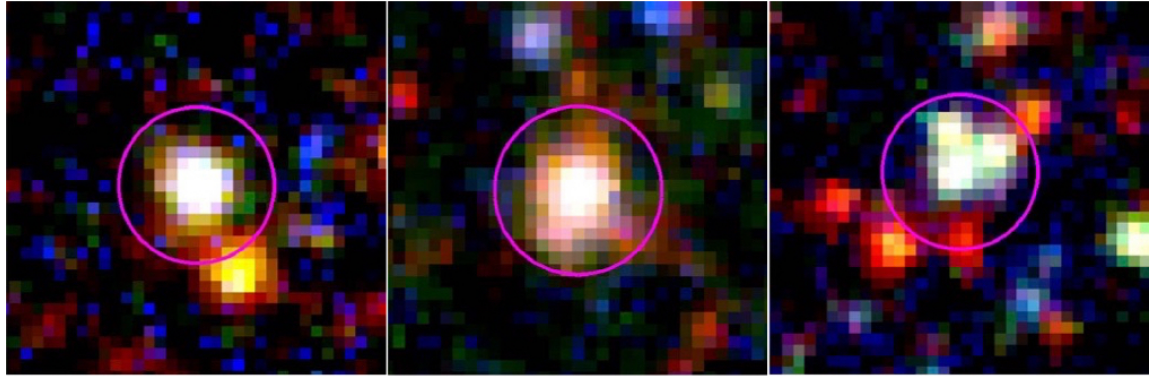
Figure 4

Examples of unresolved clusters of different morphological classes. Each panel shows a three-color UBV image of a star cluster in NGC 628, from Adamo et al. (2017). The ring shows a radius of $0.''28$, approximately 13.4 pc at the distance of NGC 628. The morphologies are classified as, from left to right, compact and symmetric, compact and asymmetric, and multiply-peaked. An "exclusive" catalog, in the sense used in this review, would include the left two objects but exclude the one on the right, while an "inclusive" catalog would include all three. A comparison to Figure 2 suggests that the ONC might well be excluded from an "exclusive" catalog.

Exclusive catalog: a star cluster catalog that excludes objects that do not pass morphological tests for compactness and roundness

Inclusive catalog: a star cluster catalog without additional morphological requirements

photometric ages for massive clusters are accurate to ≈ 0.3 dex. At ages of $\approx 5 - 50$ Myr, however, SSP colors oscillate rapidly and tracks in color space frequently cross themselves, making age determinations quite uncertain. For example, notice that SSPs with ages of ≈ 8 Myr (dark green in the figure) have very similar colors to SSPs with ages of ≈ 30 Myr (light blue in the figure), and that the separation between them is mostly along the extinction vector, making these two possibilities difficult to disentangle, particularly once stochastic color variation due to finite IMF sampling is taken into account (also see Maíz Apellániz 2009 and Krumholz et al. 2015a).

On top of these errors that apply to individual clusters, catalog construction and completeness correction induce additional errors on measurements of population demographics. Completeness correction for unresolved observations faces essentially the same challenges as for resolved data. Recently Krumholz et al. (2018) proposed a method to carry out full forward-modelling of an entire photometric catalog, simultaneously accounting for completeness and stochasticity, but this method is thus far experimental. Catalog construction is the more serious issue, because it raises the question of how one should handle sources where the light distribution is multiply-peaked or contains significant color gradients, as illustrated in Figure 4. Some authors, who limit the definition of cluster to dynamically-relaxed structures (c.f. § 1.2), discard such sources on the grounds that they are not relaxed (e.g., Bastian et al. 2011, 2012), while others adopt a broader definition and thus retain them (e.g., Chandar et al. 2010, 2011, 2014); some split the difference by including all objects but then reporting separate analyses for objects of differing morphology (e.g., Adamo et al. 2017). These different methods of catalog construction have historically led to extreme confusion in interpretation of observations, since inclusive catalogs tend to have $\approx 30 - 50\%$ more members in the youngest age bin than exclusive ones, with much smaller differences at older ages (Chandar et al. 2014; Messa et al. 2018b). In the discussion that follows, we will minimize this confusion by identifying whether a particular catalog was constructed using criteria that are "exclusive" (i.e., the catalog excludes non-symmetric objects) or "inclusive", and pointing out when the choice of one method or the other leads to systematic differences. However, we caution that all existing cluster catalogs for galaxies at distances $\gtrsim 20$ Mpc are necessarily inclusive, because limited resolution makes morphological measurement impossible in

such distant samples.

Authors who build exclusive catalogs generally assume that morphology can be used as a proxy for boundedness. This likely holds in one direction: since star-forming regions are morphologically complex, if the observed stars have a round, compact morphology, it is likely that they have relaxed into it and thus are bound. However, the converse need not be true, i.e., there is no reason to assume that a population that is too young to have dispersed is unbound simply because it is not round. The implication is that, for young stellar populations, an exclusive approach probably omits everything that is not bound, but also discards some unknown number of bound systems. By contrast, an inclusive catalog captures all structures above some luminosity threshold without regard to their dynamical state.

2.2. Mass distribution

The most basic property of a star cluster is its mass, and thus the most basic distribution for star clusters within a galaxy is the observed cluster mass function (CMF).[3] Figure 5 summarises recent measurements for the CMFs of star clusters in the disks of Local Group galaxies. All observed disk CMFs appear to be reasonably well-described by a power-law $dN/dM \propto M^{\alpha_M}$, with values of $\alpha_M = -2 \pm 0.2$.[4] A slope of $\alpha_M = -2$ corresponds to equal mass per logarithmic bin, and thus is the expected slope for a completely scale-free distribution. In normal star-forming galaxies, the mass function slope is generally measured over the range from $\approx 10^3 - 10^5 \ M_\odot$. The lower limit on this range is entirely a function of observational limitations – only in a few cases do we have cluster samples that are complete enough below $\approx 10^3 \ M_\odot$ to enable measurement of a CMF, and in those cases the data are consistent with $\alpha_M \approx -2$ down to the completeness limit. Indeed, some authors have claimed that the power law continues to masses as small as $\approx 20 \ M_\odot$, corresponding to individual massive stars (e.g., Lamb et al. 2010).

The nature of the upper limit is less clear. In most galaxies, the total number of clusters is such that very few with masses $\gtrsim 10^5 \ M_\odot$ would be expected even for a pure power-law mass function, and in galaxies with cluster populations large enough that such clusters would be expected – the Antennae, M82, and various luminous infrared galaxies (LIRGs) – they are found. Consistent with this, the luminosity of the brightest cluster in a galaxy is well correlated with the luminosity of the galaxy and its star formation rate (Larsen 2002). However, some authors report evidence that the observed CMF or luminosity function is better fit by power-law that is truncated at a mass M_c (either a hard truncation or a Schechter function, $dN/dM \propto M^{\alpha_M} e^{-M/M_c}$) than a pure power-law (e.g., Bastian 2008; Larsen 2009; Bastian et al. 2012; Adamo et al. 2015, 2017); others question the statistical robustness of this conclusion (Mok, Chandar & Fall 2018). Any truncation would need to vary systematically with some other galaxy property (e.g., the Jeans mass in the disk, § 3.1.1) in order to explain the presence of more massive clusters in more luminous galaxies.

The question of whether there is a truncation has been difficult to settle due to the challenge of deriving realistic confidence intervals from observations where the dominant errors are catalog construction and translation from photometry to cluster masses; some authors using nearly identical data sets nonetheless reach differing conclusions about whether they provide evidence for a truncation (e.g., Adamo et al. 2015 versus Sun et al. 2016 on M83), and the results can be sensitive to the statistical power of the fitting method

---

[3]Mass as used to define the CMF, and as we shall use the term throughout § 2, refers to the mass that the stellar population would have had before mass loss due to stellar evolution.

[4]It is common in the cluster literature to use the letter β rather than α_M to denote the CMF index. We use α_M, and α_T for the analogous index of the cluster age function (§ 2.3), because while there is general agreement on the usage of β, there is no equivalent agreement on the symbol used to represent the cluster age function index. On the contrary, some groups use γ for this index, while others use γ to denote a completely unrelated quantity.
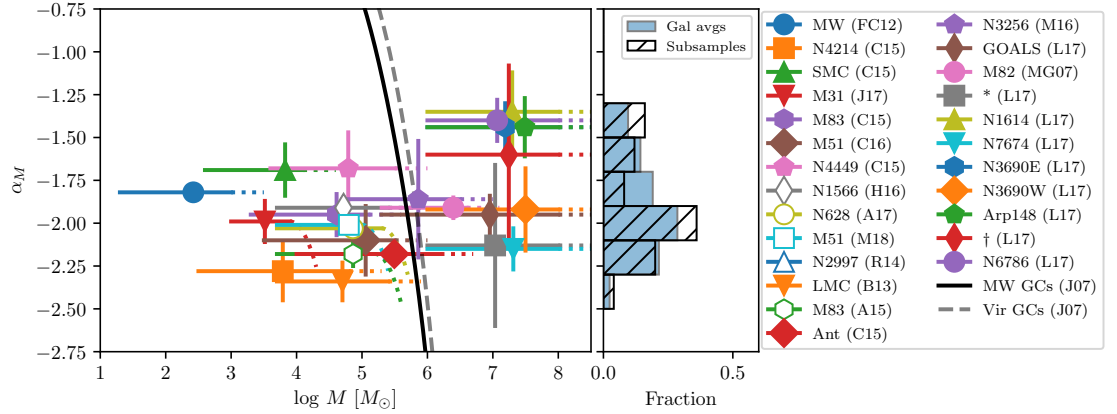
Figure 5

Compilation of measured CMF slopes α_M. In the main panel, colored points represent clusters in the disks of nearby galaxies, and solid and dashed black lines are evolved Schechter function fits to the Milky Way's GCs and to an average of GCs in Virgo, respectively. The horizontal error bars show the range in cluster mass over which the measurement was made, with open points indicating measurements from exclusive catalogs and filled points indicating inclusive catalogs. Points are displaced slightly from the centers of their corresponding mass ranges to minimize confusion. A straight dotted line on the high mass end of the horizontal error bar indicates that the observation did not detect a truncation to the CMF, while a downwards curve indicates an observation that reported a truncation. For the GC mass functions, α_M is not constant in mass, and the line shows α_M versus mass for the best-fitting evolved Schechter function. In the main panel only whole-galaxy average measurements are shown. The histogram on the right shows the distribution of measured CMF slopes for disk clusters; blue is the distribution of whole-galaxy average measurements, while hatched shows measurements in which galaxies are broken into independent sub-samples. In the legend, N is short for NGC, and the galaxies listed as ∗, †, and GOALS are UGC09618NED02, IRAS20351+2521, and an average of 22 LIRGs from the GOALS sample, respectively. References listed in the legend are as follows: MG07 = McCrady & Graham (2007), J07 = Jordán et al. (2007), FC12 = Fall & Chandar (2012), B13 = Baumgardt et al. (2013), R14 = Ryon et al. (2014), C15 = Chandar, Fall & Whitmore (2015), A15 = Adamo et al. (2015), C16 = Chandar et al. (2016), H16 = Hollyhead et al. (2016), M16 = Mulia, Chandar & Whitmore (2016), J17 = Johnson et al. (2017), A17 = Adamo et al. (2017), L17 = Linden et al. (2017), M18 = Messa et al. (2018b). The hatched histogram on the right also contains sub-samples from Messa et al. (2018a).

(Messa et al. 2018a). The most convincing case for a deviation from a pure power-law mass function is in M31, where Johnson et al. (2017) robustly identify a truncation mass of $M_c \approx 8.5 \times 10^3$ $M_\odot$ based on a sample of ≈ 1000 clusters with masses determined from CMDs rather than unresolved photometry. However, this survey does not cover some of the more actively star-forming parts of M31, and it is possible these might host clusters larger than M_c. There is also indirect evidence for a truncated CMF in the Milky Way, in the form of an observed truncation in the luminosity function for radio recombination line emission from H II regions (McKee & Williams 1997), although translating this into a truncation mass is not straightforward.

Values of α_M tend to be slightly larger for starbursting galaxies than for quiescent spirals. However, it is unclear if this represents a real physical difference or an observational artifact: the starburst galaxies are located at systematically larger distances, and thus observations of them have systematically lower spatial resolution, which will tend to flatten the mass function by blurring multiple small objects into a single large one. This matter will be "resolved" in the next decade by AO-assisted imagers and spectrographs on

extremely large telescopes.

The mass function for the GCs found in the halos of galaxies is substantially different. It has a clear turnover mass below which the number of clusters is either flat or increasing with mass. The distribution can be reasonably well described as a lognormal with a characteristic mass of $\approx 10^{5.3}\ M_\odot$ (Brodie & Strader 2006). However, a more physically-motivated description is the evolved Schechter function (Fall & Zhang 2001; Jordán et al. 2007), which is simply a standard Schechter function modified by the assumption that all clusters have lost a fixed amount of mass Δ; we discuss the physical motivation for this model in § 4.2.2. This gives a distribution $dN/dM \propto (M + \Delta)^{\alpha_{M,i}} \exp\left[-\left(M + \Delta\right)/M_c\right]$, where $\alpha_{M,i}$ is the slope at formation. The corresponding present-day slope is

$$\alpha_M = \frac{d}{d \log M}\left(\log \frac{dN}{dM}\right) = \frac{-M\left(M - \alpha_{M,i}M_c + \Delta\right)}{M_c\left(M + \Delta\right)},\tag{1}$$

In Figure 5, we show α_M for the Milky Way GC system and for the average of GC systems in the Virgo cluster from Jordán et al. (2007);[5] the best-fit M_c and Δ values are both in the range $10^{5.3} - 10^{6.6}\ M_\odot$, spanning the range of truncation masses inferred for disk clusters. However, the functional forms of the GC and disk cluster mass functions are clearly very different, and the relationship between the two is a topic to which we shall return below.

2.3. Age distribution

A second distribution of equal importance to the CMF is the age distribution of star clusters, the cluster age function (CAF). For Milky Way GCs this distribution is centred at ages ~ 10 Gyr with a width of a few Gyr (Bastian & Lardo 2018). For clusters in disks, observers generally fit the CAF as a power-law $dN/dT \propto T^{\alpha_T}$, where T is the cluster age[6]; some observers use a single α_T for all clusters, while others consider more complex forms where α_T can be different for different age or mass ranges. However, in practice most extragalactic samples cover a small enough dynamic range in mass and age that one can define at most two or three bins with potentially differing α_T values.

In a galaxy with a constant cluster formation rate, α_T is a measure of how long clusters survive. Consider a simple model, in the spirit of Boutloukos & Lamers (2003), where the probability per unit time that a cluster of age T is destroyed is $1/aT$ for some constant factor a. If we consider a cohort of clusters at a single age, our assumed destruction rate implies that their number will decrease with time as $dN/dT = -N/aT$. Integrating, we find that the number that survive to age T is $N = N_0(T/T_0)^{-1/a}$, where N_0 is the number of clusters at age T_0. This corresponds to a CAF with $\alpha_T = -1/a$. Thus $\alpha_T \to 0$ corresponds to the case $a \gg 1$, i.e., clusters survive for many times their current age, while $\alpha_T = -1$ corresponds to $a = 1$, i.e., the expected time required to destroy a cluster is equal to its current age.

While the interpretation of α_T is straightforward, divining a consistent value of it from the published literature is much more difficult. In part this is because different groups tend to subdivide the sample to be fit in different ways, and to reach divergent conclusions both about the value of α_T and whether it is different at different ranges of cluster age or mass. Figure 6 shows a summary of recent observational determinations for extragalactic systems, along with our fit to the completeness-corrected catalog of Piskunov et al. (2018)

---

[5]Jordán et al. (2007) give their fits in terms of the z-band luminosity. We have converted these values to masses using their favored z-band mass to light ratio. The line shown in Figure 5 corresponds to M_c and Δ values equal to the average of the values given in their Table 3, where $\alpha_{M,i} = -2$ in their fits. Also note that, unlike elsewhere in this section, we have not corrected for mass loss due to stellar evolution. Doing so would shift the distribution to the right by a factor of $\approx 1.5 - 2$.

[6]Throughout this review we shall use T to denote cluster ages, and t for all other time quantities.
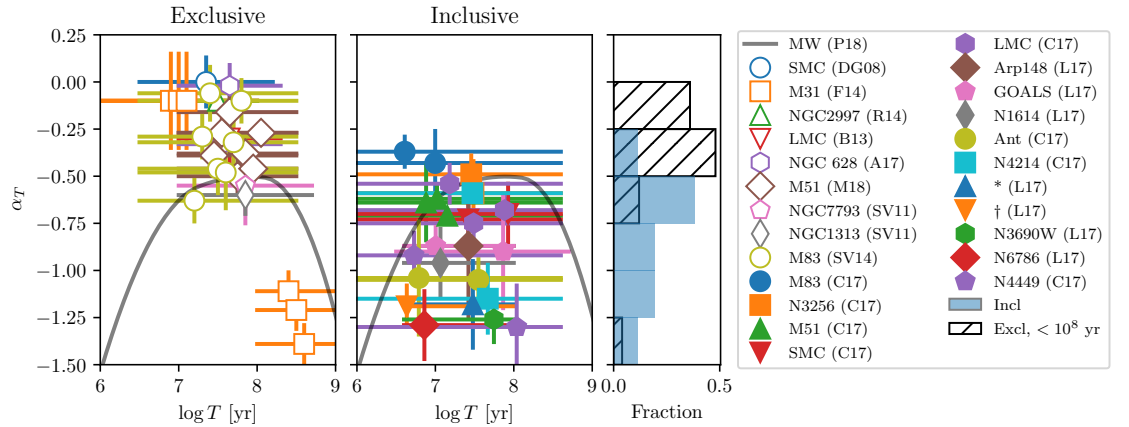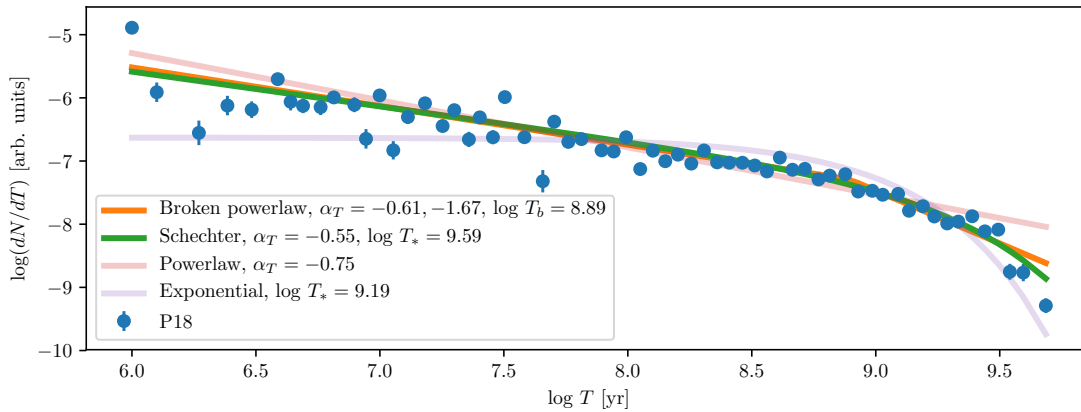
Figure 6

Observational determinations of the CAF in galaxy disks. The horizontal range indicates the age interval over which the measurement was made, while the value plotted in the vertical direction is the index α_T $(dN/dT \propto T^{\alpha_T})$ with error bars. For extragalactic data, open points correspond to exclusive catalogs (left panel), filled points to inclusive ones (middle panel). The galaxy and source are as indicated in the legend (using the same abbreviations as in Figure 5), and multiple points for a single galaxy correspond to subsamples of clusters in that galaxy, separated either by mass or by location; we have added a small random displacement to the horizontal positions of points to prevent them from overlapping. The black line shown in both the left and middle panels is a smoothed basis spline fit to the catalog of Piskunov et al. (2018) for clusters within 2 kpc in the Sun. The right panel shows a histogram of α_T values for the extragalactic catalogs, with inclusive catalogs and exclusive catalogs for ages below 10^8 yr shown separately. References are as follows: F14 = Fouesneau et al. (2014), SV14 = Silva-Villa et al. (2014), R14 = Ryon et al. (2014), C17 = Chandar et al. (2017), A17 = Adamo et al. (2017), L17 = Linden et al. (2017), M18 = Messa et al. (2018a,b), P18 = Piskunov et al. (2018).

for Milky Way clusters within ≈ 2 kpc of the Sun.[7]

Figure 6 shows that there is a systematic difference in slopes between inclusive and exclusive catalogs at young ages. The fits from exclusive catalogs at $T \lesssim 10^8$ yr have $\alpha_T \approx -0.3$ to -0.2, while the inclusive catalogs cluster around $\alpha_T \approx -1$ to -0.7. Thus exclusive catalogs suggest that clusters have long survival times compared to their current ages, while inclusive catalogs imply survival times comparable to cluster ages. In addition, although this is not shown in the figure, analysis of exclusive catalogs suggests that more massive clusters have larger values of α_T than less massive ones (e.g., Silva-Villa et al. 2014; Adamo et al. 2017; Messa et al. 2018b), implying that cluster survival time is mass-dependent. A second, more subtle dependence worth noting is that starburst galaxies seem to have a slight bias toward slopes near $\alpha_T = -1$ compared to normal spirals and dwarfs – for example, the GOALS galaxies and the Antennae all have $\alpha_T \lesssim -0.9$, while, even for inclusive catalogs, the mean slope for spirals is closer to $\alpha_T \approx -0.7$. A similar increase in α_T is seen in the centers of massive spirals compared to their outskirts (e.g., Silva-Villa et al. 2014; Messa et al. 2018a). While such an environmental dependence is in accord with some theoretical expectations (e.g., Kruijssen et al. 2011), one should be cautious in putting too much weight on the observations. The regions with steeper α_T values are on average more crowded and distant than the

---

[7]Piskunov et al. (2018) report cluster density in age bins. To derive the curve shown for α_T versus age, we computed a basis spline fit to their data and calculated the derivative from it. The source code for our calculation is provided in the Supplementary Material.
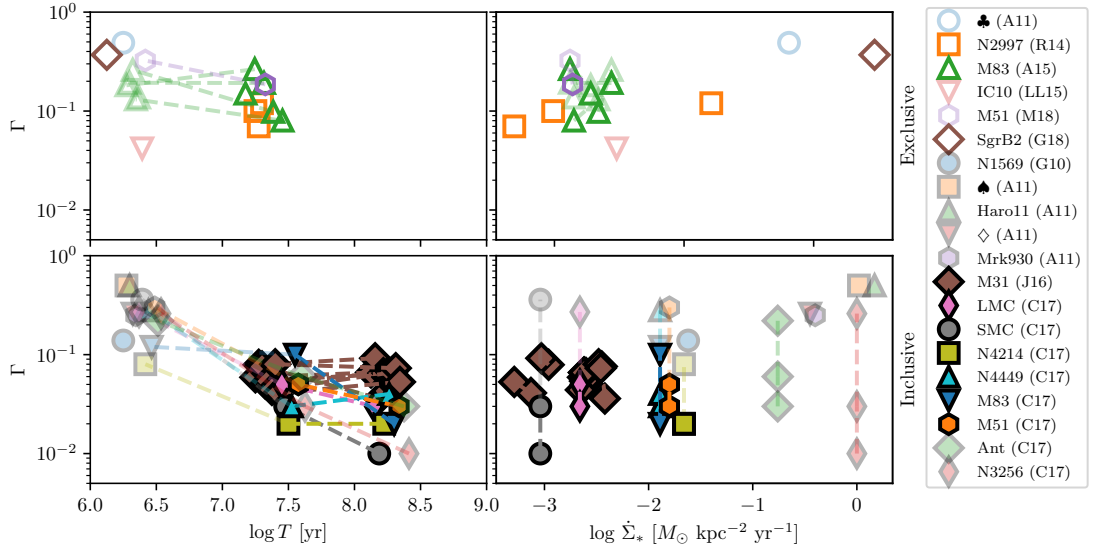
Figure 7

CAF for star clusters within ≈ 2 kpc of the Sun. Points with error bars show measurements from Piskunov et al. (2018, P18). Lines show χ^2 fits to several functional forms, with parameters indicated in the legend: broken power-law (orange; $dN/dT \propto T^{\alpha_{T,1}}$ for $\log T < \log T_b$, and $dN/dT \propto T^{\alpha_{T,2}}$ for $\log T > \log T_b$), Schechter function (green; $dN/dT \propto T^{\alpha_T} e^{-T/T_*}$), single power-law (faded red; $dN/dT \propto T^{\alpha_T}$), and exponential (faded purple; $dN/dT \propto e^{-T/T_*}$). The broken power-law and Schechter forms both have reasonable χ^2 values, while the other fits are poor.

systems with shallower α_T, and both crowding and distance could create a systematic bias in measured color that would manifest as a difference in the inferred age distribution.

Milky Way clusters show a complex dependence of α_T on age, which we highlight in Figure 7. In the Milky Way, $\alpha_T \approx -0.5$ at ages $\lesssim 10^9$ yr, indicating moderately strong cluster destruction. The age distribution steepens sharply above $\approx 10^9$ yr, indicating much more rapid disruption. This change in slope is not seen in the extragalactic samples, which are generally limited to clusters younger than $\approx 10^8$ yr for reasons of sensitivity. However, there are also two additional cautions to be mentioned in comparing the Milky Way and extragalactic samples. First, because the Milky Way sample is limited to $\lesssim 2$ kpc from the Sun, it consists entirely of low-mass clusters ($M \lesssim 10^3 \ M_\odot$), while the extragalactic sample is for much larger masses, $M \gtrsim 10^{3.5} \ M_\odot$. Second, the extragalactic sample ages are derived from photometry, while the Milky Way ages are based on CMDs, which may lead to systematic differences. CMDs are more reliable in general, but this might not hold for the Milky Way sample because many of the clusters in it contain only a small number of stars bright enough to allow placement on the CMD, leading to large age uncertainties.

2.4. Bound mass fraction

A third basic statistic for clusters is the fraction of the total stellar mass in gravitationally-bound clusters, denoted Γ. Since this fraction changes with stellar age unless $\alpha_T = 0$, Γ in general is a function of T. To measure Γ one must determine the total mass in bound clusters and the total mass in all stars within the same age interval. Measuring the former invariably involves some degree of extrapolation to account for the mass of clusters too small to detect, but for $\alpha_M \approx -2$ the extrapolation is fairly modest. For the latter quantity, at ages above ~ 10 Myr the most reliable means of determining total stellar mass is from CMDs of field stars. However, these are only available for relatively nearby sources. A second-best option is to estimate the total stellar mass by multiplying the star formation rate by the length of the age interval, assuming the star formation rate has been constant. Before proceeding, we note that several authors

Figure 8

Observational estimates of the bound mass fraction Γ. The top row shows values of Γ taken from exclusive catalogs, while the bottom row shows inclusive catalogs. In the left column we show Γ as a function of cluster age T, while in the right column we show it as a function of galaxy star formation rate $\dot{\Sigma}_*$. Measurements of Γ in multiple age ranges for the same galaxy are connected by dashed lines. Faded points indicate measurements that are potentially dubious: either measurements of Γ at ages < 10 Myr where it is not feasible to separate bound from unbound systems, or measurements of Γ for ages > 10 Myr in starburst systems where the star formation rate may have fluctuated on this time scale. In the legend, ♣, ♠, and ♢ stand for SBS0335-052E, ESO338-IG04, and ESO185-IG13, respectively. References are as follows: G10 = Goddard, Bastian & Kennicutt (2010), A11 = Adamo, Östlin & Zackrisson (2011), R14 = Ryon et al. (2014), A15 = Adamo et al. (2015), LL15 = Lim & Lee (2015), J16 = Johnson et al. (2016), C17 = Chandar et al. (2017), M18 = Messa et al. (2018b), G18 = Ginsburg & Kruijssen (2018).

describe Γ as the "cluster formation efficiency" (e.g., Larsen & Richtler 2000; Bastian 2008; Goddard, Bastian & Kennicutt 2010; Kruijssen 2012), meaning the fraction of stars formed in bound clusters. This definition implicitly assumes that one one can cleanly distinguish bound from unbound at all cluster ages, and that $\alpha_T \approx 0$ so that Γ is the same when measured for any age range. We have argued neither of these assumptions is strictly correct, but if one makes them, our more general definition of Γ reduces to theirs.

As with α_T, measurements of Γ in the literature span a wide range of values. We summarize the currently available set of measurements in Figure 8. These data must be examined with extreme caution, even in comparison to compilations of α_M and α_T. The most secure measurements, for steadily star-forming galaxies at ages above ≈ 10 Myr, generally show $\Gamma \approx 0.01 - 0.1$, consistent with the observation that the majority of early B stars are found outside clusters, as is most of the UV flux (Pellerin et al. 2007). The low value of Γ at ages $\gtrsim 10$ Myr is sometimes referred to as "infant mortality": essentially all stars are born in clusters (using our expansive definition of cluster that includes both bound and unbound systems), but a majority of the stars are not gravitationally bound, so by ages of a few tens of Myr, most stars are not members of clusters any more.

As shown in Figure 8, a number of authors have reported measurements of Γ at younger ages and in galaxies undergoing starbursts, and have attempted to deduce trends in Γ with either age (e.g., Chandar

et al. 2017) or star formation activity (e.g., Goddard, Bastian & Kennicutt 2010; Adamo, Östlin & Zackrisson 2011). We regard such claims as dubious given the methodological uncertainties, and in Figure 8 we fade the points for which significant methodological concerns exist. With regard to age trends, as pointed out by Kruijssen & Bastian (2016), estimates of Γ from inclusive catalogs at ages below ≈ 10 Myr are likely to be heavily contaminated by the presence of non-bound structures that have simply not yet had time to disperse. Conversely, however, exclusive catalogs may be missing a substantial population of bound but not yet relaxed clusters at similar ages. Thus Γ estimates at ages below 10 Myr derived from inclusive and exclusive catalogs should be viewed as upper and lower limits, respectively. An exception is the point from Ginsburg & Kruijssen (2018) for the Sgr B2 region in the Milky Way's Central Molecular Zone, for which the authors check boundedness directly using stellar velocities estimates from the radio recombination lines of the ultracompact H ii regions around each massive star.

Possible trends in Γ with star formation activity are also potentially contaminated by bias. While Γ measurements in low-star formation rate (SFR) galaxies tend to be made at ages $\gtrsim 10$ Myr to avoid the issues with inclusive versus exclusive catalogs, measurements for high-SFR galaxies are almost exclusively at younger ages. This is both because high SFR galaxies also tend to be merging systems, for which measuring Γ at older ages is problematic because the total SFR may be variable, and because high SFR galaxies are rare and thus tend to be distant, resulting in a magnitude limit that precludes measurements of Γ at ages $\gtrsim 10$ Myr (Chandar et al. 2017). This means that there is a degeneracy between two possibilities: high values of Γ in starbursts could be because Γ is in fact larger at higher SFR, but it could equally well be that Γ is higher at younger ages. The Sgr B2 point from Ginsburg & Kruijssen (2018) does not suffer from this methodological problem, but also unfortunately does not help break the degeneracy either, because it has both high SFR and young age. It is unclear to which of these factors its high value of Γ should be attributed. Consequently, we conclude that there is not yet convincing evidence that Γ varies with either age or SFR.

IMPROVING PHOTOMETRIC DETERMINATIONS OF CLUSTER PROPERTIES

The preceding sections should make clear that one of the significant uncertainties for CMFs, CAFs, and Γ are the difficulties of assigning masses and ages to clusters based on photometry. Progress will require an extensive effort "ground-truthing" photometric property determinations against measurements either CMDs or spectroscopy. Such a systematic comparison has not been performed since the work of Elson & Fall (1985, 1988) for the Magellanic Clouds using ground-based data. Much more accurate work with space-based CMDs should now be possible in both the Clouds and M31 (Johnson et al. 2016). Similarly, spectroscopy can be used to break the degeneracy between ages of $\approx 5-10$ and ≈ 50 Myr from photometry; spectroscopic measurements of a small sample of ambiguous clusters would at a minimum provide a useful prior for Bayesian analysis methods that generally return bimodal posterior PDFs when applied to clusters in the ambiguous parts of color space (Fouesneau et al. 2014; Krumholz et al. 2015a). Unfortunately the improved infrared photometry provided by the *James Webb Space Telescope* is likely to be of limited use, since clusters' colors in the IR are nearly constant at ages $\gtrsim 6$ Myr, when they become dominated by red giants and supergiants (Gazak et al. 2014).

2.5. Size and density

In addition to masses and ages, we can also study the physical sizes of star clusters, or equivalently their densities. The data on the distribution of cluster sizes are more limited than those available for mass or age, because measuring a size is obviously only possible if the object in question is at least marginally resolved. While the most physically-meaningful characterization of cluster size and density would be a half-mass radius or similar description of how mass is distributed, the limited resolution available for galaxies beyond the Milky Way and its closest neighbors generally precludes measuring anything other than the projected half-light radius, defined as the projected radius containing half the cluster light. Even this quantity is difficult to estimate for targets more distant than M31, because many clusters have relatively shallow light profiles where much of the light is in an extended halo (see § 4.1) that tends to be lost against the background. For resolved stellar populations, on the other hand, it is usually possible to extract multiple structural parameters from fits of stellar number counts to theoretically-motivated profiles; see § 4.1 and Section 1.3.2 of Portegies Zwart, McMillan & Gieles (2010) for a summary of the various size parameters that can be defined for star clusters. Here we focus on projected half-mass radii (or half-number radii if masses are not available) from the resolved samples, as these are the closest to what is available from the photometric size measurements.

Table 2   Summary of masses and radii in current cluster samples

Sample	$\log M \; [M_\odot]$[a]	$r_h \; [\mathrm{pc}]$[a]	$\log \Sigma \; [M_\odot \; \mathrm{pc}^{-2}]$[a,b]	Reference[c]
Milky Way ($d < 2$ kpc)	$1.90^{+0.51}_{-0.55}$	$0.58^{+0.32}_{-0.20}$	$1.55^{+0.31}_{-0.30}$	K13
M31 (PHAT)	$3.18^{+0.55}_{-0.55}$	$1.56^{+0.99}_{-0.46}$	$1.92^{+0.59}_{-0.45}$	J12, F14
NGC 5236 (M83)	$4.20^{+0.34}_{-0.15}$	$2.51^{+1.20}_{-0.97}$	$2.62^{+0.56}_{-0.34}$	R15
NGC 628 (LEGUS)	$4.00^{+0.40}_{-0.21}$	$3.09^{+2.16}_{-1.18}$	$2.25^{+0.56}_{-0.45}$	R17
NGC 1313 (LEGUS)	$4.10^{+0.39}_{-0.26}$	$2.69^{+1.78}_{-1.67}$	$2.51^{+0.82}_{-0.47}$	R17
M51	$4.70^{+0.30}_{-0.10}$	$3.89^{+2.99}_{-2.63}$	$2.84^{+0.84}_{-0.64}$	C16
M31 GC	$5.57^{+0.84}_{-0.70}$	$7.41^{+2.54}_{-2.11}$	$3.22^{+0.77}_{-0.96}$	B07
Milky Way GC	$5.22^{+0.42}_{-0.56}$	$3.20^{+3.88}_{-1.10}$	$3.39^{+0.60}_{-1.06}$	BH18
M82 SSC	$5.72^{+0.56}_{-0.28}$	$1.60^{+0.80}_{-0.20}$	$4.49^{+0.26}_{-0.33}$	MG07
NGC 253 SSC	$5.18^{+0.32}_{-0.86}$	$1.27^{+0.18}_{-0.32}$	$4.13^{+0.41}_{-0.66}$	L18
Milky Way YMC ($> 10^4 \; M_\odot$)	$4.39^{+0.18}_{-0.25}$	$1.50^{+2.35}_{-0.42}$	$3.21^{+0.37}_{-0.97}$	See Table 3

[a]The first number gives the sample median, while plus and minus values indicate the 16th to 84th percentile range; [b]Σ_h is the surface density at the half-mass radius, $\Sigma_h = M/2\pi r_h^2$; [c]References are as follows: K13 = Kharchenko et al. (2013), J12 = Johnson et al. (2012), F14 = Fouesneau et al. (2014), R15 = Ryon et al. (2015), R17 = Ryon et al. (2017), C16 = Chandar et al. (2016), B07 = Barmby et al. (2007), BH18 = Baumgardt & Hilker (2018), MG07 = McCrady & Graham (2007), L18 = Leroy et al. (2018).

In Figure 9 we show the mass-radius relation for cluster samples in nearby galaxies, and we summarize the properties of the cluster samples shown in the Figure in Table 2; for the Milky Way, we have updated Portegies Zwart, McMillan & Gieles (2010)'s compilation of YMCs in Table 3[8]. The plot reveals a few notable features. First, there is a mass-radius relation, but it is weak in terms of both slope and tightness. Cluster radius increases with mass slightly more slowly than $r_h \propto M^{1/3}$, so that more massive clusters have mildly higher average density. We refrain from giving a formal fit to the $\log r_h - \log M$ relation,

---

[8]Our cluster list is slightly different than Portegies Zwart, McMillan & Gieles (2010)'s compilation. Several from their list are omitted from our table because more recent observations have reduced their estimated masses below the threshold of $10^4 \; M_\odot$ commonly used to delineate YMCs, while our list includes several clusters discovered more recently than their compilation.
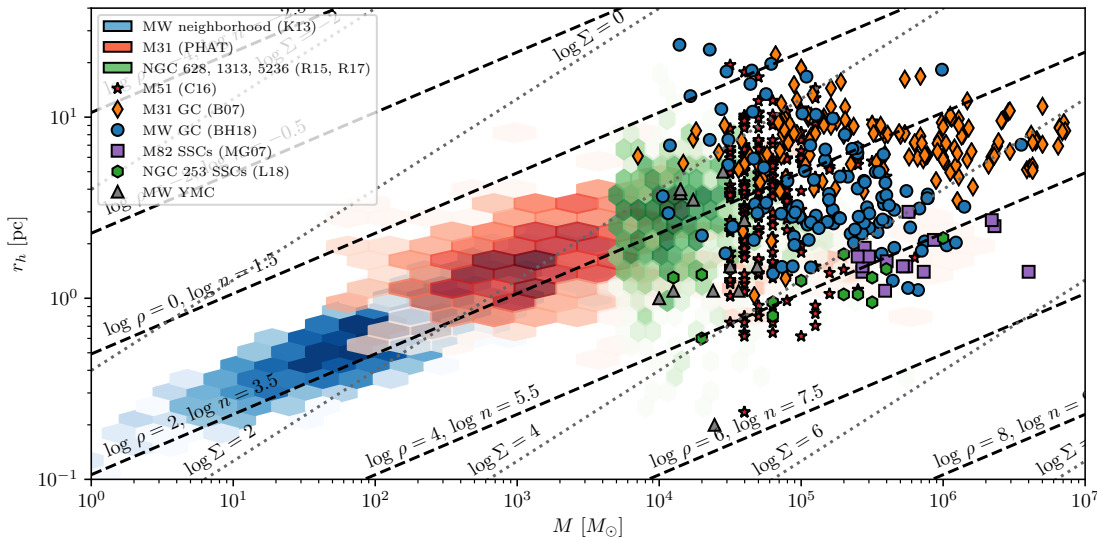
Figure 9

Mass-radius relation for star clusters in nearby galaxies. For Milky Way and globular clusters, we plot half-mass radii where available, half-number radii otherwise; for all other data sets, we plot half-light radii. Dashed black lines indicate loci of constant density within r_h, with the mass density $\rho = 3M/8\pi r_h^3$ (in units of $M_\odot$ pc^{-3}) and the corresponding number density n (units of H nuclei cm^{-3}) as indicated; dotted lines show loci of constant surface density $\Sigma = M/2\pi r_h^2$ (units of $M_\odot$ pc^{-2}). Hexagonal density plots show the log of the density of clusters in the (M, r_h) plane for Milky Way within 2 kpc of the Sun (data from Kharchenko et al. 2013, K13, with masses estimated from tidal radii using the method of Piskunov et al. 2007), clusters in M31 from PHAT (radii from Johnson et al. 2012 and masses from Fouesneau et al. 2014), and clusters in NGC 628, NGC 1313, and NGC 5236 (galaxies grouped together; Ryon et al. 2015, 2017, R15, R17). The low-mass edges in the extragalactic data are a result of observational limitations, not a physical truncation. Points show individual clusters from smaller samples: clusters in the disk of M51 (Chandar et al. 2016, C16), GCs in M31 (Barmby et al. 2007, B07) and the Milky Way (Baumgardt & Hilker 2018, BH18), SSCs in M82 (McCrady & Graham 2007, MG07) and NGC 253 (Leroy et al. 2018, L18), and YMCs from our own compilation (Table 3).

because the results would be highly sensitive to how we weighted the various data sets. Regardless of the index of the mass-radius relationship, it is clearly very broad, in the sense that, at any given mass, one can find clusters whose radii vary by a factor of 5; at least some of the spread is undoubtedly the result of measurement errors. There is no obvious difference in either the breadth or slope of the mass-radius relation between the inclusive and exclusive catalogs, although a comparison is not straightforward since, unlike for measurements of α_T and Γ, there is little overlap in Figure 9 between measurements carried out using exclusive versus inclusive catalogs. Second, the distribution is bounded at both large and small radii. Cluster densities are bounded below by $\rho \sim 1\ M_\odot$ pc^{-3} (or equivalently $n \sim 30$ H cm^{-3}); GCs in galaxy halos go to only slightly lower densities. On the small radius side, the data appear to be limited to the region $\Sigma \lesssim 10^5\ M_\odot$ pc^{-2}, as noted previously by Hopkins et al. (2010). Third, there is no separation between the loci occupied by clusters in galaxy disks and GCs; the two distributions blend continuously into one another.

Table 3   Young massive ($M > 10^4\ M_\odot$) clusters (YMCs) in the Milky Way

Name	$\log M^{\mathrm{a}}$ ($M_\odot$)	r_h^{a} (pc)	Reference[b]
Westerlund 1	4.69 ± 0.045	1.5	G11
Westerlund 2	4.56 ± 0.035	1.1	Z17
Arches	4.38 ± 0.17	1.1 ± 0.61	C12
NGC 3603	4.1 ± 0.10	1.1 ± 0.40	HEM08,R10
Quintuplet	4.0	1.0	FMM99
DBS2003 179	4.39	0.2	B08,B12
RSGC01	4.5	1.5	F06
RSGC02 (Stephenson 2)	4.6	2.7	D07
RSGC03	4.45 ± 0.15	5.0	C09
RSGC04[c] (Alicante 8)	4.15 ± 0.15	3.8	N10
RSGC05[c] (Alicante 7)	4.24 ± 0.24	3.5	N11
Alicante 10[c]	4.15 ± 0.15	4.0	GFN12

[a]Error estimates are given when provided by the referenced source; [b]References are as follows: FMM99 = Figer, McLean & Morris (1999), F06 = Figer et al. (2006), D07 = Davies et al. (2007), HEM08 = Harayama, Eisenhauer & Martins (2008), B08 = Borissova et al. (2008), C09 = Clark et al. (2009), R10 = Rochau et al. (2010), N10 = Negueruela et al. (2010), N11 = Negueruela et al. (2011), G11 = Gennaro et al. (2011), GFN12 = González-Fernández & Negueruela (2012), B12 = Borissova et al. (2012), C12 = Clarkson et al. (2012), Z17 = Zeidler et al. (2017); [c]Cited reference does not list a radius; the value we quote is a by-eye fit based on their published images, and should only be considered accurate to the factor of ~ 2 level.

2.6. Elemental abundances

In systems close enough to permit spectroscopy of individual low-mass stars, we can study the composition of clusters. The stars in all OCs in the Galaxy are observed to be highly homogeneous ($\Delta[\mathrm{Fe/H}] \lesssim 0.05$ dex) in their abundances of all elements, while no old GCs are fully homogeneous, at least not in their light elements (Li, C, N, O, F, Na, Mg, and/or Al). Essentially all massive GCs also show an anti-correlation in certain light elements (e.g., Na-O, Mg-Al; Carretta et al. 2010). Like multiple stellar populations observed in CMDs, this has become a distinguishing feature of globulars, including for globulars in the Magellanic Clouds, that is never observed in OCs (e.g., Bragaglia et al. 2017). This difference has led to the view that GC history is more complex than that of OCs (Gratton, Carretta & Bragaglia 2012). However, there is a remarkable exception to the rule: the 12-Gyr old, $10^{4.8} M_\odot$ globular Ruprecht 106 appears to be a true "single stellar population" with homogeneous abundances and no light element anti-correlations (Villanova et al. 2013); unusually, it is enriched in r-process and s-process elements but does not show the enhanced $[\alpha/\mathrm{Fe}]$ signature of the oldest stars, a characteristic shared by all other Galactic GCs. The globulars Terzan 7 and Palomar 12 also do not have multiple populations and, intriguingly, may be extragalactic through their association with the disrupting Sgr dwarf. Bragaglia et al. (2017, Table 9) catalog all modern attempts to detect multiple populations in open and globular clusters.

Typically, the heavy elements (Fe and beyond) in globulars exhibit much less scatter than for the light elements (e.g. Gratton, Carretta & Bragaglia 2012), except in a few systems (e.g., ω Cen, Terzan 5, M22, NGC 5824) that have unusual properties compared to the general family of clusters. Unusually, NGC 5824 has $\Delta[\mathrm{Fe/H}]$ variations of up to 0.3 dex but it also has a remarkable stellar halo with no tidal truncation detected out to 400 pc in radius, far beyond other clusters (Da Costa, Held & Saviane 2014). There are numerous other oddities (e.g., Marino et al. 2018): NGC 2419 has a large scatter in K-Mg; some globulars show s-process variations (e.g., M22, ω Cen), others (e.g. M15) exhibit r-process variations. With globulars, there are exceptions to every rule. In one particularly deep study of the globular NGC 6752,

Yong et al. (2013) demonstrate weak inhomogeneities in *all* elements, even the Fe peak. The existence of inhomogeneities in globulars may be related to the presence of multiple populations in these systems but there is no clear association at the present time (Bragaglia et al. 2014). Multiple stellar populations in globulars are simply not understood (Gratton, Carretta & Bragaglia 2012; Renzini 2013; Bastian & Lardo 2018).

Abundance studies in OCs are more limited. The young OC Hyades is nominally homogeneous (De Silva et al. 2006, 2011) although a deep study reveals weak variations at the level of $0.02 - 0.03$ dex (Liu et al. 2016b). The older OC M67 exhibits slightly larger variations (still $\lesssim 0.05$ dex) from a differential analysis of two fortuitous solar twins (Liu et al. 2016a). To date, all OCs have been shown to be single stellar populations with a few rare counter claims (e.g., Geisler et al. 2012) refuted in follow-up analysis (e.g., Bragaglia et al. 2014). The Hyades and M67 clusters have been subject to the most sensitive searches for abundance variations, so it is plausible that the majority of OCs will be shown to have low-level inhomogeneities in future studies. Weak metallicity variations may reflect atomic diffusion in stellar atmospheres (Dotter et al. 2017; Souto et al. 2018), planetary infall (Meléndez et al. 2017), or simply intrinsic variations that existed at the onset of star formation (Feng & Krumholz 2014). We return to this point in § 3.6.

2.7. The limits of demography at young ages

Thus far we have discussed the demographics of star clusters without too much worry about the issues raised in § 1.2 about how one defines a cluster, though our discussion about the differences between inclusive and exclusive catalogs should already hint at potential problems. We now turn directly to the question: at what stellar population age does it make sense to treat as meaningful concepts such as the CMF, CAF, and all the other demographic distributions discussed above?

To answer this question, it is helpful to conduct a thought experiment: consider an annulus within the Milky Way centered on the Solar Circle, and for every star of age $T \pm \Delta T$, find its Nth nearest neighbor (either in 3D or in projection) in the same age range, and plot the distribution of nearest neighbor distances. Qualitatively, what does this distribution look like? For stars at an age $T \sim 1$ Gyr, the answer is fairly straightforward: almost all stars at this age are field stars, and after 1 Gyr the orbits of these stars through the Galaxy are thoroughly phase-mixed. Separations therefore follow a Poisson distribution, perhaps with some overall structure at $\sim$kpc scales as a result of spiral arms. A small fraction of 1 Gyr-old stars are members of open clusters, and for these the neighbor distances are much smaller. Consequently, the overall separation distribution is a Poissonian with an excess at very small separation. Indeed, one sees exactly this sort of distribution if one plots the distribution of distances between molecular clouds and young star clusters in external galaxies (e.g., Kawamura et al. 2009). This makes it straightforward to define old clusters: they are the structures made up of the stars in the small-separation excess.

The results are very different for pre-main sequence stars (selected based on optical colours, infrared spectral energy distributions, or other methods) with ages $T \lesssim 10$ Myr. The angular correlation function for these stars is well described as a power-law $\xi(\theta) \propto \theta^{-p}$ with $p \approx 0.5 - 1.5$ on scales from ~ 0.01 pc to ~ 30 pc (e.g., Hennekemper et al. 2008; Kraus & Hillenbrand 2008; Schmeja, Gouliermis & Klessen 2009; Bressert et al. 2010). The cluster-cluster angular correlation function is a power-law up to $\sim$kpc scales (e.g., Elmegreen & Elmegreen 2001; Bastian et al. 2007; Gouliermis et al. 2017; Grasha et al. 2017a,b) – see Gouliermis (2018) for a comprehensive review. For such distributions, there is no obvious break or scale that could be used to delineate between cluster and field populations. This presents a challenge for describing the demographics of star clusters, since it is not obvious how to go about defining the clusters to begin with.

This situation leads to two main options. One is a Press-Schechter-like approach, whereby one considers the stellar field smoothed with some characteristic window that defines a size scale. However, there is an

important difference between our situation and the cosmological one: in cosmology one can compute from first principles the critical overdensity of ≈ 200 that separates bound and unbound structures, while no such calculation is available for star clusters (though there are attempts – see § 3.3). Alternately one can simply appeal to the empirical result that older clusters have a characteristic size $\sim 1 - 10$ pc, and adopt a similar averaging scale at younger ages. This is what many authors do implicitly by using clustering algorithms such as MST that use a free parameter to set the scale. A second approach is to abandon discussion of cluster demographics entirely for stellar populations $\lesssim 10$ Myr old. One can instead discuss *clustering* – i.e., the non-Poissonian distribution of stellar separations, and the correlations between separation in space and separation in other quantities (age, velocity, elemental abundance, etc.) – rather than discuss the properties of discrete clusters, an approach previously suggested by Zinnecker (2010). This removes the need to impose a size scale on a distribution where none is immediately apparent.

As stars age, their separation distribution must evolve from the power-law form seen at young ages to the Poisson plus small-separation excess form that prevails at older ages. Observations suggest that much of this evolution happens relatively quickly, over about a crossing time of the relevant size scale (e.g., Elmegreen 2000; Grasha et al. 2017b). Thus the type of demographic analysis we have discussed to this point is sensible as applied to structures more than about a crossing time old. However, from a theoretical standpoint the question that any successful theory of cluster formation must answer is how the power-law, scale-free distribution that prevails among young clusters turns into the bimodal distribution that prevails at older ages, and what determines the demographics of the clustered portion of that bimodal distribution. In other words, how does stellar *clustering* turn into star *clusters*?

3. BIRTH

In this section we examine clusters' early life, when gas dominates the mass; § 4 covers the gas-free phase. Here we first review the gaseous initial conditions for star cluster formation (§ 3.1). We then discuss observational constraints and theoretical models for the rates at which these structures form stars (§ 3.2) the stellar feedback processes that inhibit this transformation (§ 3.3), and the star formation histories that result (§ 3.4). We conclude with an attempt to synthesise this material into a coherent model for the physical properties (§ 3.5) and elemental abundances (§ 3.6) of the final gas-free cluster population.

3.1. Initial conditions for star cluster formation

3.1.1. Giant molecular clouds. Clusters are born in molecular clouds, primarily giant molecular clouds (GMCs). There is considerable evidence that GMCs obey several relations discovered by Larson (1981): (1) The 1D velocity dispersion, σ, is supersonic and varies with size L as $\sigma \propto L^p$, where $p \approx 0.5$ in the Galaxy (Solomon et al. 1987; Falgarone, Pety & Hily-Blant 2009); (2) GMCs have roughly equal kinetic and gravitational potential energies, indicating that they are bound or nearly so; (3) the column density $N \approx nL$ (or, equivalently, the mass surface density $\Sigma = M/\pi R^2$) is approximately constant within a given galaxy. As Larson pointed out, only two of these are independent. To see this, note that the degree of gravitational binding can be measured by the virial parameter, $\alpha_{\rm vir} \equiv 5\sigma^2 R/GM$, which is unity for a sphere of constant density in virial equilibrium with no surface pressure; such a sphere is bound if it has $\alpha_{\rm vir} < 2$ (Bertoldi & McKee 1992, who also considered non-spherical clouds). This definition can be rewritten as $\sigma = \sqrt{(\pi/5)G\alpha_{\rm vir}\Sigma R}$. Thus bound clouds ($\alpha_{\rm vir} \sim 1$) with the same surface density, Σ, automatically satisfy $\sigma \propto R^{1/2}$. In terms of the virial parameter, the ratio of the crossing time $t_{\rm cr} = R/\sigma$ to the free-fall time is $t_{\rm cr}/t_{\rm ff} \approx 2/\alpha_{\rm vir}^{1/2}$.

Observations of molecular gas since Larson (1981)'s work have generally confirmed his findings. For the linewidth-size relation, Solomon et al. (1987) found $\sigma = 0.72R_0^{0.5}$ km s^{-1}, where $R_0 = R/(1$ pc), in their

survey of GMCs in the first Galactic quadrant. Falgarone, Pety & Hily-Blant (2009) compiled observations of molecular gas in the Milky Way and found $\sigma \propto R^{0.5}$ within a factor 3 over the size range 0.1 to > 100 pc. Subsequent observations of Galactic (Rice et al. 2016; Miville-Deschênes, Murray & Lee 2017) and extragalactic (Bolatto et al. 2008; Faesi, Lada & Forbrich 2018) molecular gas are generally consistent with the Solomon et al. (1987) result. Faesi, Lada & Forbrich (2018) attribute the absence of a clear linewidth-size relation in some extragalactic observations to low spatial resolution. Finally, we note that the value $p \approx 0.5$ is expected for supersonic turbulence on theoretical grounds and has been seen in many simulations (e.g., Padoan et al. 2014).

One can determine if a cloud is gravitationally bound by comparing the "turbulence parameter" $C \equiv \sigma/R^{1/2}$ to the surface density Σ (Keto & Myers 1986):

$$\alpha_{\mathrm{vir}} = \frac{3.70 C^2}{\Sigma_2}, \tag{2}$$

where C is in units of km s^{-1} pc$^{-1/2}$ and $\Sigma_2 = \Sigma/(100\, M_\odot\ \mathrm{pc}^{-2})$ (c.f. Bolatto et al. 2008). The results of Solomon et al. (1987) for GMCs in the inner Galaxy ($C = 0.72$ km s^{-1} and $\Sigma_2 = 1.7$) imply $\alpha_{\mathrm{vir}} \approx 1.1$, and they concluded that these GMCs are bound. Heyer et al. (2009) and Roman-Duval et al. (2010) confirmed this, the former by plotting C vs. Σ from ^{13}CO observations of the inner Galaxy. Sun et al. (2018) were able to determine the virial parameter for the clouds in the different galaxies by measuring surface density and line width at a fixed linear scale. They found α_{vir} on the scale $R = 60$ pc between 1.3 and 3.2 in all galaxies but M31 and M33 (where they were larger), showing that the clouds are bound or nearly bound. Visual inspection of their results shows that α_{vir} is almost always close to unity for the clouds with the highest surface densities. For a cloud to be bound, it must also be able to avoid tidal disruption. For a flat rotation curve, this requires that the mean cloud density exceed twice the mean density of the galaxy inside the orbit of the cloud (Chernoff & Weinberg 1990), or $\bar{\rho} > 5.4\,(v_{c,220}/R_{\mathrm{kpc}})^2\ M_\odot\ \mathrm{pc}^{-3}$, where $v_{c,220}$ is the circular velocity in units of 220 km s^{-1}, and R_{kpc} is the distance to the center of the galaxy in kpc.

Larson's third relation, the constancy of the surface density of GMCs in the inner Galaxy (excluding the Galactic Center), was confirmed by Solomon et al. (1987), who found $\langle \Sigma \rangle = 170\, M_\odot\ \mathrm{pc}^{-2}$ and by Roman-Duval et al. (2010), who concluded that $\langle \Sigma \rangle = 140\, M_\odot\ \mathrm{pc}^{-3}$. (However, clouds with $M \lesssim 10^4\, M_\odot$ can have lower Σ – Heyer, Carpenter & Snell 2001 – and there can be systematic large-scale variations of the surface density within GMCs – Schneider et al. 2015.) Different galaxies have different surface densities: in a study of 15 disk galaxies, Sun et al. (2018) found that median surface density at 120 pc resolution ranged from about 10 to 200 $M_\odot\ \mathrm{pc}^{-2}$ excluding the Antennae, for which it is 2300 $M_\odot\ \mathrm{pc}^{-2}$. The variation in Σ for the normal galaxies could reflect variations in the beam filling factor of GMCs as well as variations in the intrinsic properties of GMCs (Lada et al. 2013). The dispersion in these values within individual galaxies is about ± 0.4 dex.

The surface density of a GMC is directly related to both its internal pressure and the pressure exerted on its surface, and insofar as Σ is constant within a galaxy, these pressures tend to be also. The turbulent pressure within a self-gravitating cloud is

$$\bar{P}_{\mathrm{turb}} \equiv \bar{\rho}\sigma^2 \equiv \frac{3\pi}{20}\alpha_{\mathrm{vir}} G\Sigma_{\mathrm{GMC}}^2, \tag{3}$$

for spherical clouds (e.g., McKee & Tan 2003), so that $\bar{P}_{\mathrm{turb}}/k_{\mathrm{B}} = 1.0 \times 10^5 \alpha_{\mathrm{vir}} \Sigma_2^2$ K cm^{-3}. This should exceed the external pressure, P_{ext}, and indeed Hughes et al. (2013) found that $\bar{P}_{\mathrm{turb}} \sim (1-8)P_{\mathrm{ext}}$ in the eight galaxies they studied. For a typical GMC in the Milky Way, with $\Sigma_{\mathrm{CO}} \sim 140\, M_\odot\ \mathrm{pc}^{-2}$, one finds that the surface pressure on the CO cloud is comparable to the mean turbulent pressure inside the cloud. On the other hand, for GMCs in high pressure environments, such as galactic nuclei or starbursts, the

surface density must be large in order for the cloud to be self-gravitating: the virial theorem implies that $\alpha_{\rm vir} \sim (1 - P_{\rm ext}/\bar{P})^{-1}$ in the absence of strong magnetic fields (Bertoldi & McKee 1992), so the requirement that $\alpha_{\rm vir} \sim 1$ implies $\Sigma \gtrsim 2(P_{\rm ext}/G)^{1/2}$. High pressures also imply high volume densities: the mean density in a cloud is

$$\bar{\rho}_{\rm GMC} = \frac{3}{4}\left(\frac{\pi\Sigma_2^3}{M_6}\right)^{1/2} \quad M_\odot \ \ \text{pc}^{-3} = \frac{179}{\alpha_{\rm vir}^{3/4}M_6^{1/2}}\left(\frac{\bar{P}_{\rm turb}/k_{\rm B}}{10^8 \ \text{K cm}^{-3}}\right)^{3/4} \quad M_\odot \ \ \text{pc}^{-3} \tag{4}$$

where the final step follows because for a non-spherical cloud $\Sigma \approx M/V^{2/3}$ has the value as for a spherical cloud of the same volume, V, and $P_{\rm turb} \gtrsim 2P_{\rm ext}$ for bound clouds.

3.1.2. Clumps and clusters. Clusters are born in density concentrations (clumps) within GMCs. Both the GMCs and the clumps within them are supersonically turbulent. In the absence of self-gravity and for an approximately isothermal equation of state, this turbulence leads to a log-normal density distribution (e.g., McKee & Ostriker 2007) and a corresponding log-normal distribution of surface densities (Brunt, Federrath & Price 2010). Self-gravity leads to the formation of a power-law tail in the PDF for dense gas (Collins et al. 2012). This has been observed in the PDF of surface densities in a number of GMCs, $dP/d\log\Sigma \propto \Sigma^{-(2\pm0.7)}$ for clouds with active star formation; this is often associated with an overall power-law density distribution in the cloud (Lombardi, Alves & Lada 2015; Schneider et al. 2015). The clumps in GMCs are formed by the combined action of supersonic turbulence and self-gravity. Equation 2 implies that the typical surface density of a bound clump within a bound GMC (each of which therefore has $\alpha_{\rm vir} \approx 1$) is about the same as that of the GMC, or $\Sigma_{\rm clump} \sim 100\,M_\odot \ \text{pc}^{-2}$ in the Galaxy. Observational studies of clumps have been carried out only for the Galaxy. Analyzing the results of the APEX 870 μm survey of the Galactic plane (excluding the central $\pm 5°$ in longitude), which was complete for $\Sigma > 700\,M_\odot \ \text{pc}^{-2}$ and $M > 1000\,M_\odot$ within 20 kpc, Urquhart et al. (2018) found about $10^7\,M_\odot$ in 8000 dense clumps. A by-eye fit to their data on the surface densities of the clumps gives $dN_{\rm clump}/d\log\Sigma \propto \Sigma^{-1.6}$ over the range $\Sigma = 600 - 6000\,M_\odot \ \text{pc}^{-2}$; the median surface density is about $700\,M_\odot \ \text{pc}^{-2}$ and the maximum is about $3 \times 10^4\,M_\odot \ \text{pc}^{-2}$. The clumps are gravitationally bound with $\alpha_{\rm vir} \sim 0.1 - 1$; Kauffmann, Pillai & Goldsmith (2013) and Tan et al. (2013) have suggested that magnetic fields could provide support for sub-virial clumps. At high masses, the clump mass distribution has $\alpha_{M,\rm clump} \approx -2$ (indeed, Heithausen et al. 1998 found $\alpha_{M,\rm clump} = -1.84$ over 5 decades in mass) and has an upper limit of about $10^5\,M_\odot$. In the Galaxy, the GMCs out of which the clumps form have a power-law mass with $\alpha_{M,\rm GMC} \approx -1.6$, so that most of the mass of GMCs is in massive clouds. This distribution is truncated at about $10^7\,M_\odot$ (Rice et al. 2016 and references therein, but see Miville-Deschênes, Murray & Lee 2017 for a different view), as expected theoretically from the Jeans mass of the galactic disk (i.e., the Jeans mass expressed in terms of the gas surface density, $\Sigma_{\rm gal}$; Kim & Ostriker 2001),

$$M_J = \frac{\sigma^4}{G^2\Sigma_{\rm gal}} = 10^7\left(\frac{\sigma}{7 \ \text{km s}^{-1}}\right)^4\left(\frac{13\,M_\odot \ \text{pc}^{-2}}{\Sigma_{\rm gal}}\right) \quad M_\odot. \tag{5}$$

If the Toomre Q-parameter, $Q = \kappa\sigma/(\pi G\Sigma) \sim 1$, where κ is the epicyclic frequency, this is about the maximum mass expected for a bound cloud, so in the Galaxy the most massive dense clump is about 1% of the mass of the most massive GMC.

Observations of GMCs in external galaxies show that in some, but not all, cases the mass distribution can be described by a truncated power law with most of the mass in the most massive clouds, as observed in the Galaxy: this describes the GMCs in NGC 300 (Faesi, Lada & Forbrich 2018) and those in the inner region and in the spiral density waves of M51, but not the GMCs in the rest of the galaxy, where $\alpha_{M,\rm GMC} < -2$, so that low-mass clouds contain most of the mass (Colombo et al. 2014). In the regions

of M51 in which $\alpha_{M,\,\rm GMC} > -2$, the upper limit on the mass distribution is similar to the Galactic value, $\sim 10^7\,M_\odot$.

The conditions required to produce massive clusters, such as globular clusters, occur in interacting galaxies such as the Antennae (e.g., Whitmore et al. 2014), dwarf irregular galaxies, some of which are starbursts and many of which show signs of interaction (Billett, Hunter & Elmegreen 2002), and nuclear starbursts in spiral galaxies such as NGC 253 (e.g., Leroy et al. 2018). Such conditions are rare now, but were common at redshifts $z \gtrsim 2$ (Kruijssen 2014). A detailed survey of the GMCs in the Antennae galaxies has yet to be made, but Wilson et al. (2000) carried out a low-resolution ($\sim 300 \times 500$ pc) survey of this system and identified five large molecular cloud complexes with masses in excess of $10^8\,M_\odot$; the largest one not associated with one of the galactic nuclei had a mass of $6 \times 10^8\,M_\odot$. By contrast, Wilson et al. (2003) pointed out that the most massive molecular cloud complex in M51 had a mass about 10 times less. They suggested that the interaction of the galaxies produced regions of less shear, which enables the existence of more massive GMCs; Billett, Hunter & Elmegreen (2002) emphasized the importance of low shear in producing more massive clusters in dwarf irregular galaxies. For disk galaxies, we can express the Jeans mass (Equation 5) in terms of the Toomre Q-parameter, so that

$$M_J = \frac{\pi^4 G^2 \Sigma^3 Q^4}{\kappa^4} = 2500 \frac{\Sigma_2^3 Q_{1.5}^4}{\Omega_0^4}\; M_\odot, \tag{6}$$

where in the second expression a flat rotation curve has been assumed, Q has been normalized to a typical critical value, and Ω_0 is measured in units of Myr^{-1} (Krumholz & McKee 2005). Hence, if Q self-regulates to be of order unity, a reduction in the epicyclic frequency, κ, by a factor 2 due to the effect of the interaction can increase the maximum GMC mass by more than an order of magnitude. More massive GMCs enable the production of more massive clumps (Harris & Pudritz 1994; Reina-Campos & Kruijssen 2017); Johnson et al. (2015a) have observed a clump in the Antennae with a mass of at least $5 \times 10^6\,M_\odot$ and a radius less than 24 pc, corresponding to $\Sigma \gtrsim 3000\,M_\odot\,\mathrm{pc}^{-2}$ and $\bar{P}_{\rm turb}/k_B \gtrsim 10^8$ K cm^{-3}.

The maximum possible mass of a cloud that forms by gravitational instability in a disk is $M_{\rm GMC,\,max} \simeq (\lambda_{\rm T}/2)^2 \Sigma$, where $\lambda_{\rm T} = 4\pi^2 G\Sigma/\kappa^2$ is the Toomre length (Escala & Larson 2008). Disks are stable for $Q \gtrsim 1.5$ (Kim & Ostriker 2001), and the Jeans mass (eq. 6) at $Q = 1.5$ is close to $M_{\rm GMC,\,max}$. If the disk has a total mass (including stars and dark matter) inside a radius R of $M_{\rm tot}(< R)$, a gas mass inside R of $M_g(< R) \simeq \pi R^2 \Sigma$ and it is rotationally supported, then

$$\frac{M_{\rm GMC,\,max}}{M_g(< R)} \simeq \frac{4\pi}{(\kappa/\Omega)^4}\left[\frac{M_g(< R)}{M_{\rm tot}(< R)}\right]^2 \quad : \tag{7}$$

the maximum fraction of the gas that can go into a single GMC scales as the square of the gas fraction. As a result, more massive GMCs, and therefore more massive cluster-forming clumps, are expected in gas-rich regions such as galaxies in the process of formation and in starbursts (Escala & Larson 2008).

Galaxy interactions also produce shocks that can compress clouds. A collision at a relative velocity $v_{\rm rel}$ produces radiative shocks with a velocity $v_{\rm rel}/2$ and a pressure $P = \rho_0(v_{\rm rel}/2)^2$, corresponding to $P/k_B = 3.8 \times 10^6 n_{\rm H,0}(v_{\rm rel}/300\,\mathrm{km\,s}^{-1})^2$. This leads to localized regions of very high pressure when clouds collide, $P/k_B \sim 10^8$ K cm^{-3} (Jog & Solomon 1992), but only $\sim 10^6$ K cm^{-3} for shocks in the intercloud medium, which is not that much larger than the typical pressure in Galactic GMCs.

Based on this summary of the observed properties of molecular gas in galaxies, we have the following expectations for the gaseous precursors of star clusters currently forming in the Galaxy: clusters form from clumps with masses ranging from very small values (depending on the star formation efficiency) up to about $10^5\,M_\odot$ and surface densities from $\sim 100\,M_\odot\,\mathrm{pc}^{-2}$ up to about $3 \times 10^4\,M_\odot\,\mathrm{pc}^{-2}$. A substantial fraction of the clouds are bound ($\alpha_{\rm vir} \lesssim 2$). The typical clump has a column density comparable to that of the

cloud in which it is embedded. For a clump mass of $10^3\, M_\odot$ and a typical GMC pressure corresponding to $\Sigma_{\rm GMC} = 140\, M_\odot\ {\rm pc}^{-2}$, namely $P_{\rm turb}/k_{\rm B} \sim 2\times10^5$ K cm^{-3}, the density is about $50\, M_\odot\ {\rm pc}^{-3}$. Clumps have a distribution of surface densities extending to higher values, and the mean density of a clump at a given mass scales as $\Sigma_{\rm clump}^{3/2}$ (Equation 4). Conditions in nearby spiral galaxies do not differ substantially from those in the Milky Way. These conditions are consistent with the production of the OCs observed in the local universe, but not of GCs, which are far more massive than the clumps observed in the Galaxy. Molecular clouds that can produce globulars are seen in interacting galaxies like the Antennae, dwarf starbursts such as He 2-10 (Johnson et al. 2018), and nuclear starbursts (Leroy et al. 2018), and are expected in any disk with a high gas fraction (as observed in galaxies at high redshift – Genzel et al. 2011) or a small epicyclic frequency (as observed in interacting and dwarf irregular galaxies).

3.2. Conversion of gas to stars

The collapse of molecular clouds eventually produces stars. Here we review only those aspects of this process most relevant to star clusters; for a more general treatment see the reviews by McKee & Ostriker (2007) and Krumholz (2014). The CMF and the production of bound clusters depend strongly on the efficiency with which gas is converted into stars. Efficiency can be defined in multiple ways, which we can illustrate in an idealized example: a cloud of initial mass M and free-fall time $t_{\rm ff}$ forms stars at a rate $\dot{M}_*$. At any given time, the remaining gas mass is M_g and the mass of stars formed is M_*, and we can parametrize the relationship between the star formation rate, current gas mass, and free-fall time as

$$\dot{M}_* = \epsilon_{\rm ff}\frac{M_g}{t_{\rm ff}}, \tag{8}$$

where $\epsilon_{\rm ff}$ is one of the efficiencies we will define. A region in free-fall collapse with nothing inhibiting star formation has $\epsilon_{\rm ff} \approx 1$. In addition to gas consumption by star formation, feedback from stars launches a wind at a rate $\dot{M}_w$, which we normalize to the star formation rate by defining the mass loading factor $\eta = \dot{M}_w/\dot{M}_*$. With these definitions, we have

$$\dot{M}_g = -\,(1+\eta)\,\epsilon_{\rm ff}\,\frac{M_g}{t_{\rm ff}}. \tag{9}$$

All the factors that appear in this equation – η, $\epsilon_{\rm ff}$, or $t_{\rm ff}$ – can depend on M_g, M_*, time, or any number of other variables. Moreover, real clouds are not closed boxes with a fixed initial mass; there is certainly mass inflow occurring simultaneously with star formation (see "Conveyor-Belt Model for Cluster Formation" in the Supplementary Materials). Indeed, nothing in our formulation even requires that clouds be bound – M_g may can include both bound and unbound material. For the purposes of illustration, however, we can ignore these complications and treat η, $\epsilon_{\rm ff}$, and $t_{\rm ff}$ as constant, in which case it is trivial to write down expressions for the instantaneous gas and stellar masses:

$$\frac{M_g}{M} = e^{-t/t_{\rm sf}} \qquad \frac{M_*}{M} = \frac{1 - e^{-t/t_{\rm sf}}}{1+\eta} \equiv \epsilon_*\left(1 - e^{-t/t_{\rm sf}}\right) \qquad t_{\rm sf} \equiv \frac{\epsilon_*}{\epsilon_{\rm ff}}t_{\rm ff}. \tag{10}$$

These equations provide another definition of the star formation efficiency: ϵ_*, the fraction of the initial cloud mass that has been transformed into stars once all the gas is consumed. In this idealized problem, $\epsilon_* = 1/(1+\eta)$. The term $t_{\rm sf}$ we have introduced defines the characteristic timescale for star formation, i.e., M_*/M reaches its final value of ϵ_* on a timescale $t_{\rm sf}$. In this section we discuss $\epsilon_{\rm ff}$, and in the next two sections we examine ϵ_* and $t_{\rm sf}$.

3.2.1. Observational constraints on $\epsilon_{\rm ff}$. The value of $\epsilon_{\rm ff}$ is well constrained by observations. Formally, for any specified region of volume V containing a gas mass M_g, we have $\epsilon_{\rm ff} = \dot{M}_*/[M_g/t_{\rm ff}(\bar{\rho})]$, where $\dot{M}_*$ is the instantaneous star formation rate within V, $t_{\rm ff}(\bar{\rho}) = \sqrt{3\pi/32G\bar{\rho}}$ is the free-fall time as a function of density, and $\bar{\rho} = M_g/V$ is the mean density of the region in question. There are several methods available to estimate these quantities. The most direct is to define a column density or extinction threshold, estimate the mass within that threshold, calculate the density and thus the free-fall time from the mass and projected area (assuming the unseen third dimension is comparable in size to the two observed dimensions), and estimate the star formation rate by counting young stellar objects (YSOs) and estimating their masses and the duration of the YSO phase (Krumholz, Dekel & McKee 2012; Federrath 2013; Evans, Heiderman & Vutisalchavakul 2014; Salim, Federrath & Kewley 2015; Heyer et al. 2016; Ochsendorf et al. 2017; also see Heiderman et al. 2010; Lada et al. 2013). All published studies using this method rely on data from *Spitzer*, which was only sensitive enough to detect $\sim 1\ M_\odot$ YSOs within a few kpc of the Sun, and massive YSOs out to the distance of the Magellanic Clouds. Consequently, this method is only available out to the Magellanic Clouds, and studies that go beyond the Solar neighborhood must make a correction for the unseen part of the IMF.

A second approach, usable throughout the Milky Way and its satellites, is to match catalogs of star-forming regions identified by tracers such as infrared or free-free emission with catalogs of molecular clouds identified by CO or dust emission, matching them up if they are sufficiently close in position or velocity, and then using the mass, free-fall time, and star formation rates of the matched clouds and star-forming regions to estimate $\epsilon_{\rm ff}$ (Vutisalchavakul, Evans & Heyer 2016; Lee, Miville-Deschênes & Murray 2016; Ochsendorf et al. 2017). A third approach, available for extragalactic systems with extensive molecular gas and star formation tracer maps, is simply to pixelize the entire galaxy, and estimate masses, densities, and free-fall times in each pixel (Krumholz, Dekel & McKee 2012; Leroy et al. 2017; Utomo et al. 2018). A fourth method is to observe tracers of dense gas, most commonly HCN, and correlate these with tracers of star formation (e.g., Krumholz & Tan 2007; García-Burillo et al. 2012; Hopkins et al. 2013; Usero et al. 2015; Gallagher et al. 2018). For this method there is no need to spatially resolve the emission, because the molecule itself determines the density and free-fall time; e.g., Onus, Krumholz & Federrath (2018) show that the luminosity-weighted mean density of HCN line-emitting gas is $\approx 10^4$ cm^{-3}.[9]

We summarize recent observational constraints on $\epsilon_{\rm ff}$ in Figure 10. Comparing the various methods, we see that essentially all studies based on YSO counting or HCN give $\epsilon_{\rm ff} \approx 0.01$, with a study-to-study dispersion of ≈ 0.3 dex, and a dispersion of about $0.3 - 0.5$ dex within any single study. This dispersion is probably an upper limit, since it includes both the true physical dispersion and the effects of measurement errors. Pixel statistics produce a similar dispersion, but with a median $0.3 - 0.5$ dex lower. However, these results are still consistent because there are systematic uncertainties in all the methods at the ~ 0.5 dex level. For pixel statistics one must assume a value for the poorly-constrained line of sight depth through the target galaxy. For YSO counting, one must estimate the duration of the phase during which newly-formed stars would be classified as YSOs, the volume density of the gas seen only in projection, and the IMF correction. For HCN, there is uncertainty in the mean density of the emitting gas and the emissivity per unit mass, which is needed to estimate the total mass.[10]

The cloud matching studies disagree with the other three methods, yielding much larger dispersions

---

[9]A fifth and final method is available for the special case of the Milky Way's Central Molecular Zone (CMZ), where one can use the positions of clouds and star clusters around their orbit through the CMZ as an absolute clock to time the transformation of gas into stars (Barnes et al. 2017). Since this method is applicable only to the CMZ we will not discuss it further.

[10]Figure 10 uses the Onus, Krumholz & Federrath (2018) calibration, which is intermediate between the alternative Kauffmann et al. (2017) and Gao & Solomon (2004) estimates.
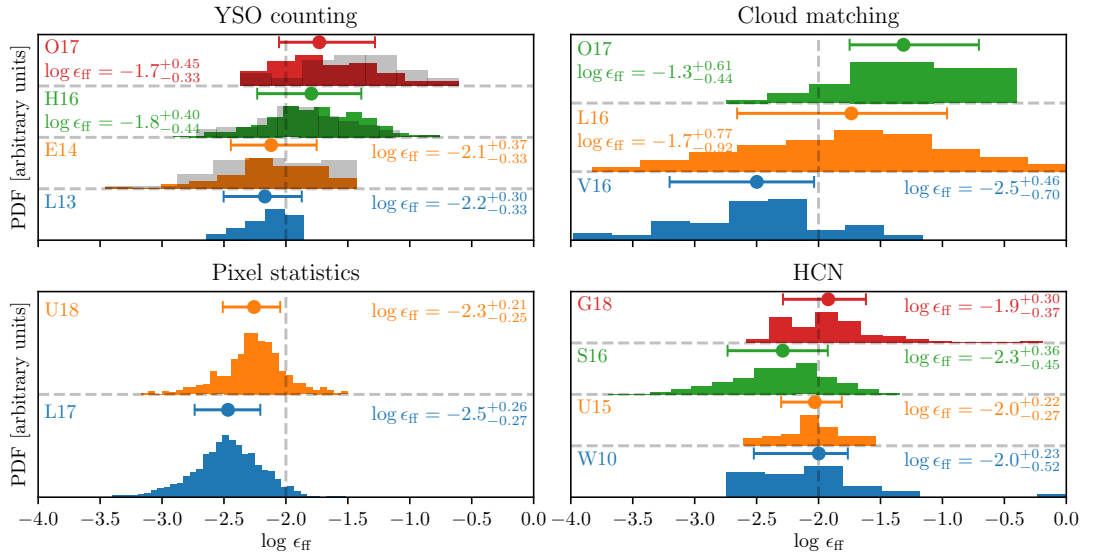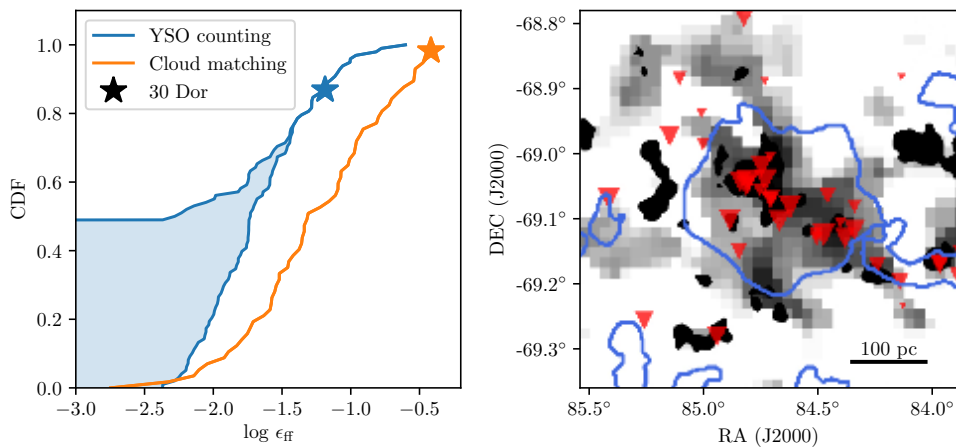
Figure 10

Compilation of $\epsilon_{\rm ff}$ measurements, grouped by method. Each histogram shows the distribution of $\epsilon_{\rm ff}$ values measured in a single study. Points and error bars show the median and the 16th to 84th percentile; numerical values are printed in each panel. In panels with two histograms, the colored one is the distribution including non-detections plotted at their 2σ upper limits, while the gray histogram shows the distribution omitting non-detections; median and percentile values are for the colored distribution. The vertical dashed lines at $\log\epsilon_{\rm ff} = -2$ are to help guide the eye. References are as follows: W10 = Wu et al. (2010), L13 = Lada et al. (2013), E14 = Evans, Heiderman & Vutisalchavakul (2014), U15 = Usero et al. (2015), H16 = Heyer et al. (2016), L16 = Lee, Miville-Deschênes & Murray (2016), V16 = Vutisalchavakul, Evans & Heyer (2016), S16 = Stephens et al. (2016), L17 = Leroy et al. (2017), O17 = Ochsendorf et al. (2017, appears twice, because this study used two independent methods), G18 = Gallagher et al. (2018), U18 = Utomo et al. (2018). All values of $\epsilon_{\rm ff}$ are as reported by the authors of the study, except that we have derived $\epsilon_{\rm ff}$ for the Lada et al. (2013) sample following Krumholz (2014), and we have homogenized $\epsilon_{\rm ff}$ from the HCN data sets to the calibration of Onus, Krumholz & Federrath (2018).

and mostly higher medians. The main difference between cloud matching and the other methods is that cloud matching does not require that the star formation and gas tracers be co-spatial, which makes the results sensitive how one constructs and matches catalogs of clouds and star-forming regions; the 0.8 dex difference in median $\epsilon_{\rm ff}$ values derived by Vutisalchavakul, Evans & Heyer (2016) and Lee, Miville-Deschênes & Murray (2016) is almost entirely due to this sensitivity, since the regions they target and the underlying data sets they use are nearly identical. To gain more insight into why cloud matching differs from other methods, it is helpful to examine the study of Ochsendorf et al. (2017), who use both YSO counting and cloud matching to derive $\epsilon_{\rm ff}$ values for the same molecular gas map of the LMC.[11] For YSO counting, they adopt the standard approach of counting YSOs within the same contours used to define clouds, while for cloud matching they decompose an Hα map into H II regions and match clouds and H II regions if there is any spatial overlap. We show the distribution of $\epsilon_{\rm ff}$ values produced by the two methods in the left panel

---

[11]Ochsendorf et al. (2017) analyze both a dust-based molecular cloud map from Jameson et al. (2016) and a CO-based map from Wong et al. (2011). We focus on the dust-based map because, due to the LMC's low metallicity, is contains substantial amounts of CO-dark molecular mass.

Figure 11

Left: cumulative distribution of ϵ_{ff} values measured in the study of Ochsendorf et al. (2017) by YSO counting versus cloud matching, both using the Jameson et al. (2016) H_2 map. The star shows the 30 Doradus region. For the YSO counting method, the blue band indicates the uncertainty range associated with non-detections, with the lower value corresponding to assuming that clouds without YSO detections have star formation rates at the sensitivity limit, and the upper value corresponding to assuming that such regions have star formation rates of zero. *Right:* zoom-in on the 30 Doradus region, reproduced from Ochsendorf et al. (2017) by permission of the AAS. Grayscale shows the H_2 column density, black contours show CO clouds from the catalog of Wong et al. (2011), red triangles show the positions of massive YSOs, with size indicating inferred mass (Ochsendorf et al. 2016), and blue contours show H II regions.

of Figure 11.

Clearly cloud matching yields a broader distribution with a higher median. To understand why, we can look at the example of the 30 Doradus region (indicated by the star in the left panel of Figure 11) for which YSO counting gives $\epsilon_{\mathrm{ff}} = 0.065$, while cloud matching gives $\epsilon_{\mathrm{ff}} = 0.38$, a factor of 6 larger. Since the same GMC properties are used in both estimates, the difference arises solely from the imputed star formation rate. Examining the right panel of Figure 11, it is clear how a substantial difference could arise: while the great majority of the YSOs are within the footprint of the cloud, there is significant $H\alpha$ emission that is not, but that partly overlaps it. Treating all this emission as if it comes from the cloud causes the cloud matching method to infer a much larger star formation rate. This emission could indeed be associated with the existing molecular cloud, but it could also be associated with molecular gas that has been dispersed or displaced by stellar feedback.

Similar effects are also apparent in other galaxies. For example, in a survey of NGC 628 at ~ 50 pc resolution, Kreckel et al. (2018) identify ≈ 1500 H II regions and ≈ 750 GMCs, but find < 100 overlaps between the two. Naive application of a cloud matching method to this data set would lead one to conclude that most GMCs are inert and have $\epsilon_{\mathrm{ff}} = 0$, while most H II regions formed with $\epsilon_{\mathrm{ff}} \gg 1$, an obviously unphysical result. The data are better explained by the hypothesis that bright H II regions rapidly disperse or displace the molecular clouds in which they form. In principle it should be possible to account for this effect and extract corrected estimates of ϵ_{ff} from the statistics of the GMC-H II region correlation (Kruijssen et al. 2018), but this method has not yet been widely applied.

This analysis suggests that cloud matching-based estimates of ϵ_{ff} should be treated with caution. Given

THE FUTURE OF YSO COUNTING

YSO counting is the most reliable method of estimating $\epsilon_{\rm ff}$, but at present it is limited by the sensitivity and angular resolution of *Spitzer*, the telescope used for all existing YSO counting studies. With *Spitzer*, the only YSOs we can reliably detect beyond the Milky Way have masses $\gtrsim 8\ M_\odot$ (Ochsendorf et al. 2016). This sensitivity limit is responsible for the large uncertainty in the shape of the low end of $\epsilon_{\rm ff}$ distribution shown in Figure 11. However, the *James Webb Space Telescope* will be far more sensitive and have significantly better angular resolution. With *JWST* it will possible to detect $\approx 1\ M_\odot$ YSOs in the Magellanic Clouds, and to obtain YSO catalogs comparable to those currently available for the LMC and SMC out to M31 at least, possibly farther. This will greatly improve cloud-scale measurements of $\epsilon_{\rm ff}$ beyond the Milky Way.

the reasonable consistency in results produced by the other methods, we conclude that the preponderance of the current observational evidence favors $\epsilon_{\rm ff} \approx 0.01$ for regions $\gtrsim 1$ pc in size, with a dispersion of $\lesssim 0.5$ dex and a comparable level of systematic uncertainty.

3.2.2. Theory of $\epsilon_{\rm ff}$. The observed value of $\epsilon_{\rm ff}$ is surprisingly low, and a number of authors have proposed theoretical models aimed at explaining it. The earliest models relied on magnetic regulation of star formation. Interstellar gas is magnetized, and if the field strong enough it can prevent gas from collapsing until non-ideal magnetohydrodynamic effects reduce the magnetic flux threading the gas. However, more recent measurements of magnetic fields have generally shown that they are too weak to support the gas – see Crutcher (2012) for a review. Consequently, attention has focused on three possibilities: $\epsilon_{\rm ff}$ may be low because clouds are not bound, because they are turbulent, or due to feedback.

Unbound cloud models propose that $\epsilon_{\rm ff}$ is low because most of the material observationally-defined as molecular clouds is not in fact self-gravitating (e.g., Dobbs, Burkert & Pringle 2011; Meidt et al. 2018). We have argued in § 3.1.1 that the bulk of the evidence does not support this view, but if it were the case that only $\sim 1\%$ of the mass in a molecular cloud were bound, this would naturally explain why $\epsilon_{\rm ff} \approx 1\%$. While this is a potentially-viable explanation for the low $\epsilon_{\rm ff}$ values of GMC as traced by CO, most of the observations compiled in § 3.2.1 are for the much denser gas traced by HCN, or by the YSO counting studies, which is almost certainly bound (e.g., Kauffmann, Pillai & Goldsmith 2013). The unbound cloud hypothesis does not explain why this gas also shows low $\epsilon_{\rm ff}$.

Turbulent regulation models propose that $\epsilon_{\rm ff}$ is low because star-forming gas is turbulent enough to render most sub-regions of a molecular cloud unbound, even if that cloud is bound on its largest scales. Qualitatively the argument stems from the linewidth-size relation discussed in § 3.1: since velocity dispersion varies with size as $\sigma \propto R^{1/2}$, then the kinetic energy per unit mass contained within a region of size R scales as $e_K \propto R$, while the binding energy per unit mass of a region of mass M scales as $e_G \propto GM/R$. If one chooses a region close to the mean density then $M \propto R^3$ and $e_G \propto R^2$. Consequently, for a randomly-chosen sub-region of a cloud, the virial ratio obeys $\alpha_{\rm vir} \propto 1/R$, and thus most regions smaller than an entire cloud are unbound. This argument can be made quantitative by integrating over the density PDF. The earliest version of such an argument appeared in Krumholz & McKee (2005), but this has been superseded by numerous later works that model the density PDF and the evolution of its self-gravitating parts with increasing accuracy (Padoan & Nordlund 2011; Hennebelle & Chabrier 2011; Padoan, Haugbølle & Nordlund 2012; Federrath & Klessen 2012; Hopkins 2013; Burkhart 2018; Burkhart & Mocz 2018).

The consensus of recent models is that turbulence does substantially reduce $\epsilon_{\rm ff}$, but not all the way to $\epsilon_{\rm ff} \approx 0.01$ as required by the observations. As time goes by the density PDF of a self-gravitating cloud

deviates develops an increasingly-prominent power-law tail on its high-density end that causes ϵ_{ff} to rise with time (Murray & Chang 2015; Burkhart 2018). This density build-up is most likely counteracted by localized feedback processes that break up high-density regions and suppress the growth of the power-law tail; the most likely mechanism for this is protostellar outflows, since stars begin to launch outflows as soon as they form, and even low-mass stars can produce significant outflow feedback. Modern simulations that include both protostellar outflows and magnetic fields[12] that help couple them to gas are able to achieve values of ϵ_{ff} that, while still a factor of $2-3$ too high compared to the consensus observational value, are approaching agreement with the data (Wang et al. 2010; Myers et al. 2014; Federrath 2015; Cunningham et al. 2018). An alternative possibility is that cloud disruption by feedback is so fast and efficient that typical clouds never have time to develop power-law tails extensive enough to drive ϵ_{ff} to large values (Grudić et al. 2018).

A final caveat on theoretical explanations for the low value of ϵ_{ff} is that turbulent regulation is effective only if the gas is actually turbulent. There is significant debate in the literature about whether the observed large linewidths of molecular clouds should be interpreted as turbulence, or whether they might represent coherent gravitational collapse (e.g., Heitsch, Ballesteros-Paredes & Hartmann 2009; Ballesteros-Paredes et al. 2011; Zamora-Avilés, Vázquez-Semadeni & Colín 2012; Traficante et al. 2018). While there is significant evidence from both analytic theory (Klessen & Hennebelle 2010; Goldbaum et al. 2011) and numerical simulations (Robertson & Goldreich 2012; Lee & Hennebelle 2016a,b; Birnboim, Federrath & Krumholz 2018) that accretion flows invariably drive turbulence, and recent kinematic measurements with *Gaia* also appear to disfavor the idea that GMCs are in global collapse (§ 3.4.2), the question is not settled.

3.3. Feedback and the termination of star formation

The final star formation efficiency, ϵ_*, depends both on ϵ_{ff} and on the ability of stellar feedback to eject gas from nascent star clusters. Unfortunately we cannot measure ϵ_* for individual clouds directly, because for a single region we cannot measure both the gas mass at the onset of star formation and the stellar mass at its conclusion. One can attempt to measure the mass-loading factor η, but this is technically very challenging (see Yang et al. 2018 for a recent attempt). The best prospect for measuring ϵ_* is likely the use of statistical methods to analyze populations of clouds (Kruijssen et al. 2018), but thus far measurements using this technique are not widely available. For these reasons, estimates of ϵ_* come primarily from theory. In the remainder of this section we follow the approach taken by several previous authors (e.g., Matzner 2002; Fall, Krumholz & Matzner 2010; Matzner & Jumper 2015; Rahner et al. 2017) by considering a variety of feedback mechanisms and attempting to determine under what circumstances they become relevant. We summarize these findings in Figure 12.

3.3.1. Protostellar outflows. Bally (2016) provides a comprehensive recent review of protostellar outflows, so we focus only on the details most relevant for star cluster formation. Outflows are critical to breaking up dense regions and thus keeping ϵ_{ff} small because they eject $\sim 2/3$ of the mass from individual protostellar cores (e.g., Matzner & McKee 2000; Offner & Chaban 2017). However, this mass is ejected at relatively low velocities, and thus may not escape the larger-scale protocluster. For this reason, theoretical models (e.g., Matzner & McKee 2000; Matzner & Jumper 2015), numerical simulations (e.g., Wang et al. 2010; Krumholz, Klein & McKee 2012; Murray, Goyal & Chang 2018), and observations (e.g., Nakamura & Li 2014; Li et al. 2015; Plunkett et al. 2015) all suggest that outflows have a limited role in setting final

---

[12]Non-MHD simulations find that outflows are much less effective at lowering ϵ_{ff}, since without magnetic fields outflows couple poorly to the bulk of the material (e.g., Krumholz, Klein & McKee 2012; Murray, Goyal & Chang 2018).
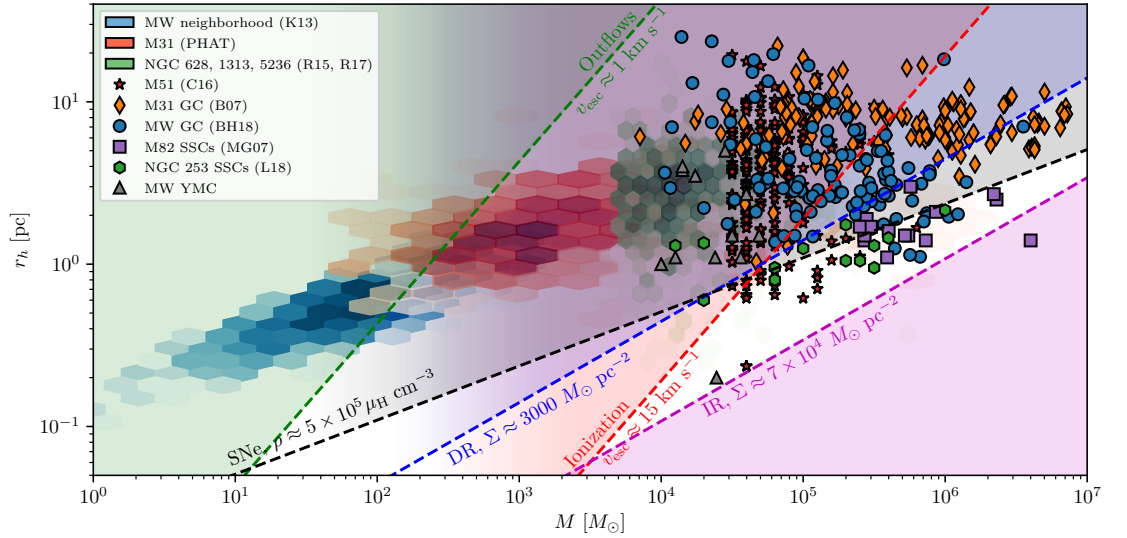
Figure 12

Star cluster mass-radius relation, showing the same data as Figure 9, with shaded regions indicating where various feedback mechanisms are potentially significant. In constructing this diagram, we have for simplicity considered the case of star-forming clouds where the gaseous and stellar components have comparable masses and radii, but we emphasize that in reality this need not be the case. The feedback mechanisms shown are outflows (§ 3.3.1), supernovae (SNe; § 3.3.6), direct radiation pressure (DR; § 3.3.3), ionization (§ 3.3.2), and indirect radiation pressure (IR; § 3.3.4). Shaded regions fade at masses below which stochastic sampling of the IMF makes them unlikely to be active (§ 3.3.1). For all feedback types except IR, the region where the mechanism is effective is to the left of and above the line.

star formation efficiencies in most clusters. There has yet to be a comprehensive numerical survey of the parameter space, but Matzner & Jumper (2015) estimate analytically that outflows produce $\epsilon_* \lesssim 0.5$ only in clusters with escape speeds $v_{\rm esc} \lesssim 1$ km s^{-1}. We show this limit in Figure 12.

Despite this limitation, outflows play a unique role, because they do not depend on the presence of massive stars. The momentum per unit mass of stars formed delivered by protostellar outflows to their surroundings is of order the escape speed from a protostellar surface (e.g., Matzner & McKee 2000). Low-mass and high-mass protostars have similar surface escape speeds, so low-mass stars are as effective at providing outflow feedback as massive ones. By contrast, for all the other mechanisms we discuss below, feedback is dominated by massive stars unlikely to be found in small clusters. To quantify this point, we use the SLUG stochastic stellar populations code (da Silva, Fumagalli & Krumholz 2012; Krumholz et al. 2015b) to measure the PDFs for the number of supernovae, and bolometric and ionizing luminosity, in clusters at a range of masses for a Chabrier (2005) IMF. We find that the number of expected supernovae exceeds unity only for clusters with masses $\gtrsim 100$ $M_\odot$, and that the median (bolometric, ionizing) luminosity of clusters is $< 50\%$ of the value for a fully-sampled IMF in clusters with mass $\lesssim (400, 700)$ $M_\odot$. Consequently, protostellar outflows are likely to be the only feedback mechanism that limits ϵ_* for stellar populations smaller than a few hundred $M_\odot$. This limit is very blurry due to stochasticity; some clusters with masses of only a few hundred $M_\odot$ nevertheless have substantial ionizing luminosities (Andrews et al. 2014).

3.3.2. Photoionization feedback.

A population of stars that fully samples the IMF produces ionizing photons at a rate of $\Psi_i \approx 6 \times 10^{46}$ s^{-1} $M_\odot^{-1}$ (Murray & Rahman 2010). These can heat gas in star-forming clouds to $\approx 10^4$ K, giving it a sound speed of ≈ 10 km s^{-1}. If not trapped either by the surrounding material or by gravity, this gas will flow outward in a wind known as a champagne flow, directly removing mass from the cloud. The back-reaction created when the ionized gas pushes off the surrounding neutral material, known as the rocket effect, may eject even more mass and exert significant forces, ultimately unbinding clouds entirely. The mass removed in this process can be considerable – the analytic calculations of Williams & McKee (1997) and Matzner (2002) suggest that a 10^4 $M_\odot$ population of stars can eject ≈ 10 times its own mass over a few Myr, giving $\epsilon_* \sim 0.1$. Numerical simulations generally confirm this result, though they show that the value of ϵ_* produced by ionization also depends strongly on the initial boundedness of clouds, and on whether or not they are magnetized (e.g., Dale, Ercolano & Bonnell 2012, 2013; Howard, Pudritz & Harris 2017; Geen et al. 2015, 2016; Gavagnin et al. 2017; Dale 2017; Kim, Kim & Ostriker 2018).

The main limitation on photoionization as a feedback mechanism is that the ionized gas must be able to escape, but there is some tension in the simulation literature about how much of a limitation this represents. Dale, Ercolano & Bonnell (2012, 2013) find that photoionization is unable to unbind more than a few percent of the mass for clouds with $v_{\rm esc} > 10$ km s^{-1}, while Kim, Kim & Ostriker (2018) find $\epsilon_* \lesssim 0.5$ at all escape speeds $\lesssim 20$ km s^{-1}, though it is possible that this is because they include direct radiation pressure as well as photoionization (see § 3.3.3). None of the other published studies explore a range of $v_{\rm esc}$ values. It is plausible that even regions with $v_{\rm esc} > 10$ km s^{-1} could lose mass due to photoionization, since freshly-ionized gas rocketing off a dense neutral surface will generally accelerate to small multiples of the ionized gas sound speed. In the analogous problem of photoionized gas driving winds off accretion disks, simulations show that even regions with escape speeds ≈ 30 km s^{-1} launch significant winds (Woods et al. 1996). Thus escape speed required to suppress the effects of photoionization is of order 10 km s^{-1}, but its exact value is uncertain; in Figure 12 we adopt 15 km s^{-1}.

3.3.3. Direct radiation pressure.

A zero-age population of stars that fully samples the IMF has a light to mass ratio $\Psi \approx 1100$ $L_\odot$ $M_\odot^{-1}$ (e.g., Fall, Krumholz & Matzner 2010). The light is emitted mostly at UV wavelengths where the ISM is highly opaque, so that even a modest dust column is sufficient to ensure that essentially all the radiation momentum is deposited in the gas. Whether this affects ϵ_* depends on how the associated radiation force compares to pressure and gravitational forces. H II region gas pressures vary with radius as $r^{-3/2}$ due to ionization balance, while direct radiation pressure varies as r^{-2}, so gas pressure dominates once H II regions reach a characteristic size $r_{\rm ch} \approx 0.03 S_{49}$ pc, where S_{49} is the ionizing luminosity in units of 10^{49} s^{-1} (Krumholz & Matzner 2009). For a stellar population of mass M_* with ionizing luminosity $S = \Psi_i M_*$, one can show that this radius is larger than the size of the cluster, and thus radiation pressure rather than photoionization dominates mass removal, only if the escape speed $v_{\rm esc} \gtrsim 10$ km s^{-1}. Thus direct radiation pressure becomes dominant roughly where photoionized gas pressure ceases to be effective.

The importance of direct radiation pressure relative to gravity depends on the cloud column density (Fall, Krumholz & Matzner 2010; Murray, Quataert & Thompson 2010). Consider a spherical gas cloud of total (gas plus stellar) mass M and radius R, with a stellar population of mass $\epsilon_* M$ forming at its center. The column-averaged outward force per unit gas mass supplied by radiation is $\Psi \epsilon_* M / 4\pi R^2 (1-\epsilon_*)\Sigma c$, where $\Sigma = M/\pi R^2$ is the mass per unit area. The corresponding column-averaged inward force per unit mass from gravity is $\approx GM/R^2$. Thus the Eddington ratio, which defines the ratio of radiative to gravitational

force, is $f_{\text{Edd}} = \Sigma_{\text{DR}}\epsilon_*/(1 - \epsilon_*)$, where

$$\Sigma_{\text{DR}} = \frac{\Psi}{4\pi Gc} \approx 340 \, M_\odot \ \text{pc}^{-2} \tag{11}$$

is the surface density below which radiation direct pressure becomes important. Naively we would expect direct radiation pressure to expel gas and thus limit ϵ_* only for $f_{\text{Edd}} \gtrsim 1$, requiring clouds with $\Sigma \lesssim \Sigma_{\text{DR}}$. However, Thompson & Krumholz (2016) point out that in a turbulent medium there will be sightlines with $\Sigma < \Sigma_{\text{DR}}$ even if $\Sigma > \Sigma_{\text{DR}}$ on average. Gas can be ejected on these low-column sightlines, leading to a wind. Their models suggest that, in a cloud with $\epsilon_{\text{ff}} \approx 0.01$, this wind will remove mass fast enough to produce $\epsilon_* \approx 0.5$ at $\Sigma \approx 10\Sigma_{\text{DR}}$; we show this line in Figure 12.

Simulations of the effects of direct radiation pressure have produced mixed results. All simulators find that direct radiation pressure is ineffective at reducing ϵ_* at surface densities $\gtrsim 10\Sigma_{\text{DR}}$ (e.g., Grudić et al. 2018), but at the surface densities of $\Sigma \approx 100 \, M_\odot \ \text{pc}^{-2}$ characteristic of Milky Way GMCs some simulators find that direct radiation pressure yields $\epsilon_* \lesssim 0.1$ (Grudić et al. 2018), while others find substantially larger values of $\approx 0.3 - 0.5$ (Raskutti, Ostriker & Skinner 2016; Kim, Kim & Ostriker 2018; Howard, Pudritz & Harris 2018). These disagreements may be due to differences in the initial conditions, or to resolution problems in the simulations that can lead to a significant overestimate in the effectiveness of direct radiation pressure in some numerical methods (Krumholz 2018).

3.3.4. Indirect radiation pressure. When interstellar dust absorbs radiation, it re-emits the energy at IR wavelengths. IR dust opacities are a factor of ~ 100 smaller than UV ones, but regions of sufficiently high column density may nonetheless be opaque to the re-radiated IR. In this case the IR photons will be absorbed again, and can deposit additional momentum; in clouds opaque enough to require many cycles of absorption and re-emission before the energy escapes, the resulting force will greatly exceed the direct radiation force (Thompson, Quataert & Murray 2005; Murray, Quataert & Thompson 2010; Thompson et al. 2015).

Indirect (i.e., dust-reprocessed) radiation pressure feedback is governed by two processes: frequency diffusion and radiation Rayleigh-Taylor (RRT) instability. Frequency diffusion is important because IR wavelengths are much larger than interstellar grain radii, so dust opacity varies with frequency as $\kappa \propto \nu^2$. Thus as stellar radiation diffuses outward through a dusty envelope and mean photon frequency decreases, dust-radiation coupling weakens. Eventually photons shift to frequencies low enough to escape. This effect greatly limits the importance of dust-reprocessed radiation (Wolfire & Cassinelli 1987; Reissl et al. 2018), and means that the mass-weighted mean opacity of a cloud is a function of the total gas column. Many subgrid models of radiation pressure ignore this effect and simply assume a constant IR opacity (e.g. Rosdahl & Teyssier 2015; Tsang & Milosavljević 2018; Hopkins et al. 2018); given the importance of diffusion, their results should be treated with caution.

RRT instability (Jacquet & Krumholz 2011) causes gas that is being accelerated out of a gravity well by indirect radiation to become clumpy, reducing the effectiveness of its coupling to radiation. Early numerical work suggested this would prevent gas ejection entirely (Krumholz & Thompson 2012), but this result proved to be sensitive to the details of the radiation transfer method. Simulations based on flux-limited diffusion (Krumholz & Thompson 2012) or M1 closures (Rosdahl & Teyssier 2015; Skinner & Ostriker 2015) find that RRT prevents gas ejection unless the dust opacity is greatly boosted compared to observed Milky Way values, while those based on variable Eddington tensor (Davis et al. 2014) or implicit Monte Carlo (Tsang & Milosavljević 2015) find that it does not, though it does greatly reduce the ejection velocity. The latter two methods are likely more reliable, so we expect that indirect radiation pressure can limit ϵ_*.

Under what conditions will this happen? Krumholz & Thompson (2012) show, and simulations by Davis et al. (2014) confirm, that for any column of gas confined by gravity, there is a maximum radiation flux

that can pass through it without triggering gas ejection. Calculations properly accounting for frequency diffusion show that a gas cloud forming stars at its center will exceed the critical flux before the star formation efficiency ϵ_* reaches 0.5 only if its surface density exceeds (Crocker et al. 2018a,b)

$$\Sigma_* \gtrsim \Sigma_{*,\mathrm{IR}} = \frac{16\sqrt{\pi G \sigma_{\mathrm{SB}}/c}}{\Psi} \left(\frac{\kappa_0}{T_0^2}\right)^{-1} \approx 7 \times 10^4 \left(\frac{\kappa_0}{0.03 \text{ cm}^2 \text{ g}^{-1}}\right)^{-1} M_\odot \text{ pc}^{-2}, \tag{12}$$

where $T_0 = 10$ K and κ_0 is the opacity per unit gas mass to radiation with a color temperature T_0; the normalization used above is typical of Milky Way dust. We show this condition in Figure 12. Note that indirect radiation pressure has a *minimum* surface density at which it becomes effective, while all the feedback mechanisms discussed previously have a *maximum* surface density (for direct radiation) or escape velocity (for ionization) below which they are effective. The reason for this difference is the frequency diffusion effect: for a fixed radiation flux, a larger gas column traps heat more effectively and thus the amount of indirect radiation force delivered rises superlinearly with the gas column. Consequently, indirect radiation force becomes more effective relative to gravity as the total column density increases.

3.3.5. Hot star winds. Stars with surface temperature $\gtrsim 25{,}000$ K and metallicity $\gtrsim 0.5 Z_\odot$ drive winds with speeds of several thousand km s^{-1} (e.g., Vink, de Koter & Lamers 2001), producing temperatures $\gtrsim 10^7$ K when the wind shocks against the ISM. At this temperature the radiative cooling time is long compared to the dynamical time (unlike for protostellar outflows – § 3.3.1 – or winds from intermediate-mass stars, Offner & Arce 2015), leading to the formation of hot bubbles that can push on surrounding colder gas and potentially eject it from a cloud. There are a number of analytic models describing this process in one dimension without cooling (see Bisnovatyi-Kogan & Silich 1995 for a review) and with cooling (Koo & McKee 1992), and several authors have published 1D numerical results including cooling (e.g., Krause et al. 2016; Fierlinger et al. 2016; Rahner et al. 2017; Silich & Tenorio-Tagle 2018; Naiman, Ramirez-Ruiz & Lin 2018). The general conclusion from these models is that winds will be the dominant feedback mechanism unless their effects are reduced by two phenomena that cannot be properly captured in 1D: hydrodynamic leakage of hot gas through low-density channels, and sapping of hot gas thermal energy via turbulent mixing at the hot-cold interface. The general conclusion from these models is that winds will be the dominant feedback mechanism unless their effects are reduced by two phenomena that cannot be properly captured in 1D: hydrodynamic leakage of hot gas through low-density channels, and sapping of hot gas thermal energy via turbulent mixing at the hot-cold interface.

To constrain the importance of leakage and mixing, and thus the effectiveness of hot star winds at reducing ϵ_*, there are two approaches available: multi-dimensional simulations and observations. Published simulations find that, as a result of efficient leakage (Rogers & Pittard 2013; Calura et al. 2015; Wareing, Pittard & Falle 2017) or mixing-induced losses (Mackey et al. 2015), wind feedback lowers ϵ_* by at most tens of percent. However, there has yet to be a comprehensive parameter study. In addition, only Wareing, Pittard & Falle (2017) consider magnetized clouds, and the fields they assume are weak. It is possible that more realistic, stronger magnetic fields might reduce mixing losses (Gentry et al. 2018). Moreover, winds may be important not so much because they remove mass, but because they halt ongoing accretion of low-density gas onto molecular clouds (e.g., Gatto et al. 2017; Haid et al. 2018).

X-ray observations probe wind feedback because the pressure in hot gas is directly related to its X-ray luminosity. The observed luminosities of diffuse gas around YMCs is rule out the possibility that the wind is trapped (Dunne et al. 2003; Townsley et al. 2003, 2006, 2011), and require significant leakage or mixing-induced radiative loss (Harper-Clark & Murray 2009; Rosen et al. 2014; Rogers & Pittard 2014; Toalá & Arthur 2018). Indirect diagnostics based on IR line ratios are consistent with this conclusion (Yeh & Matzner 2012). Comparing the hot gas pressure in H II regions to the pressure exerted by warm ionized

gas and radiation, Lopez et al. (2011, 2014) find that hot gas is subdominant in the massive 30 Doradus region, and in a wide range of smaller H ɪɪ regions in the LMC and SMC. Pellegrini, Baldwin & Ferland (2011) reach the opposite conclusion for 30 Doradus, but only under the assumption that the hot gas has a small volume filling factor, which Lopez et al. (2014) point out would make it locally significant but unimportant as an agent of gas clearing on large scales. Thus there is currently no observational evidence that hot star winds are important in setting ϵ_*. However, as with the simulations, observations cover only a narrow range of parameter space, and it is possible that future work might identify regimes where winds are important.

3.3.6. Supernovae. Supernovae (SNe) deliver about the same energy as hot star winds, but all at once rather than gradually over several Myr. Whether this makes them more effective is uncertain. Some simulators find that SNe effectively destroy molecular clouds (Rogers & Pittard 2013; Calura et al. 2015; Wareing, Pittard & Falle 2017), while others find that their effects are modest compared to other feedback mechanisms (Körtgen et al. 2016; Geen et al. 2016; Rey-Raposo et al. 2017). Simulation results appear to be sensitive to both resolution and numerical method (Gentry et al. 2017, 2018), and there has yet to be a systematic survey of parameter space. Moreover, analytic models suggest that photoionization feedback prior to the first SN strongly affects the final outcome (McKee, van Buren & Lazareff 1984; Matzner 2002), but of the published studies, only Geen et al. (2016) include this effect.

One important limitation on the importance of SN feedback is time delay: even the most massive stars require $t_{\rm SN} \approx 3$ Myr to explode, and sufficiently dense regions may convert a significant fraction of their gas to stars before the first SN occurs (Fall, Krumholz & Matzner 2010; Kruijssen 2012). A region with free-fall time $t_{\rm ff}$ and star formation rate efficiency $\epsilon_{\rm ff}$ and no feedback mechanisms except SNe will convert a fraction $\epsilon_* = \epsilon_{\rm ff} t_{\rm SN}/t_{\rm ff}$ of its mass to stars before the first SN. This yields $\epsilon_* < 0.5$ only if the density obeys

$$\rho < \rho_{\rm SN} \approx \frac{3\pi}{128 G \epsilon_{\rm ff}^2 t_{\rm SN}^2} \approx 5 \times 10^5 \mu_{\rm H} \left(\frac{\epsilon_{\rm ff}}{0.01} \times \frac{t_{\rm SN}}{4 \text{ Myr}} \right)^{-2} \text{ cm}^{-3}, \tag{13}$$

where $\mu_{\rm H} = 2.3 \times 10^{-24}$ g is the mean mass per H nucleus for the usual cosmic composition of 27% He by mass. We show this criterion in Figure 12. Moreover, SNe may be unimportant even in regions that obey Equation 13 because other mechanisms might remove gas on a timescale below $t_{\rm SN}$. There is substantial observational evidence that this is the case in at least some regions: clusters with ages $\approx t_{\rm SN}$ are observed to be gas free both in the Milky Way (Longmore et al. 2014) and in M83 (Hollyhead et al. 2015), and the spectra of clusters in the process of gas clearing frequently show Wolf-Rayet features, indicating an evolutionary phase that precedes the first SN (Sokal et al. 2016). These observations strongly disfavor SNe as an important mechanism for gas removal. However, much like winds, the main effect of SNe may not be to clear dense gas, but rather to remove low-density gas and thereby prevent star formation from restarting after gas clearing.

3.3.7. Summary of feedback mechanisms and expectations on ϵ_*. Figure 12 provides an overall picture of stellar feedback. For the smallest clusters, $M \lesssim 100 \ M_\odot$, protostellar outflows are the only mechanism of interest because the massive stars responsible for other feedback mechanisms are likely to be absent. Moving to slightly higher masses, in the great majority of the parameter space occupied by observed clusters in the disks of modern galaxies, and even for a substantial portion of the globular cluster population, SNe, photoionization, and direct radiation pressure are all at least potentially effective. When all three are capable of operating, both simulations and observations suggest that photoionization will dominate because it has no time delay, and photoionized gas pressure exceeds radiation pressure except in very compact star clusters with large escape speeds. Most star formation in this region should be characterized by efficiencies

$\epsilon_* \lesssim 0.3$, since simulations suggest that photoionization feedback is quite effective when it operates. This is also in accord with Lada & Lada (2003)'s observation that, for embedded clusters with masses $\gtrsim 10^3\ M_\odot$, the stellar mass fraction seems to increase with age, but only to a maximum of $\approx 30\%$.

As we move to yet higher masses and densities, characteristic of globular clusters, YMCs, and SSCs, these three feedback mechanisms begin to become ineffective (see also Grudić et al. 2018). Ionization likely becomes ineffective first, followed by direct radiation pressure and SNe; this ordering is dictated simply by the mechanisms' dependence on mass and radius, as shown in Figure 12. Presumably ϵ_* increases in this region. Indirect radiation, unlike all the other feedback mechanisms, only becomes effective once the surface density rises above a threshold value. The observed cluster mass-radius distribution shows a maximum surface density (first pointed out by Hopkins et al. 2010) consistent with this limit, suggesting that indirect radiation pressure either lowers ϵ_* at surface densities above the threshold, or causes cloud radii to expand during star formation to drive them below it (Crocker et al. 2018a).

Interestingly there is a locus, represented by the unshaded triangle in Figure 12, where no known feedback mechanism is expected to be effective. This locus is occupied by a small population of massive, dense star clusters, generally classified as SSCs. Our analysis suggests that these clusters may form with very high ϵ_* values, simply because no feedback mechanism is capable of ejecting much gas from them. However, we caution that it is probably not realistic to assume that protoclusters spend their entire formation history on this locus; as we discuss below, real protoclusters almost certainly grow by gas accretion and form stars simultaneously, and thus will move in Figure 12.

3.4. Star formation histories and timescales

3.4.1. Observational constraints. The final ingredient in our simple model of star cluster formation is its duration $t_{\rm sf}$, or, more generally, the star formation history. Age-dating stars is a fraught problem (Soderblom et al. 2014), so the conclusions we draw must be regarded as somewhat tentative, though the situation has improved considerably in the last few years.

By far the most comprehensively-studied region is the Orion complex. The young stars in Orion are spread over several tens of pc, but molecular gas and ongoing star formation are limited to a smaller ~ 10 pc region around the ONC, which is ≈ 1 pc in size (Hillenbrand & Hartmann 1998 find $r_h \approx 0.8$ pc) and is centered on θ^1 Ori C, the most massive star in the complex. The ONC is the densest region, and, despite its small volume, represents a significant fraction of the total number of stars in the complex. We illustrate one recent measurement of the age distribution in Orion in Figure 13, which suggests three important conclusions that appear to apply to other star-forming regions as well (e.g., Tan, Krumholz & McKee 2006; Azimlu, Martínez-Galarza & Muench 2015; Getman et al. 2018; Caldwell & Chang 2018). First, the densest regions show age spreads that are substantially larger than their dynamical times. Using the ages shown in Figure 13, the time over which 90% of the ONC stars formed is ≈ 6 Myr (other studies give similar results: Reggiani et al. 2011; Da Rio et al. 2016; Beccari et al. 2017), while the free-fall time is ≈ 0.6 Myr (Tan, Krumholz & McKee 2006; Da Rio, Tan & Jaehnig 2014). Consequently, star formation has been ongoing for ≈ 10 free-fall times. While the absolute ages and the magnitude of the age spread depend on the choice of age indicator, the conclusion that there is a substantial age spread is robust against this choice, and against possible contamination due to photometric variability (Messina, Parihar & Distefano 2017) or variable accretion histories (see Sidebar). This conclusion is also bolstered by kinematic observations showing that the ONC is very close to virial equilibrium (Kim et al. 2019), implying that it must have had time to relax dynamically.

Second, while star formation has an extended history, the star formation rate has not been constant over this period; star formation has accelerated (Palla & Stahler 2000; Caldwell & Chang 2018). (In Figure 13, recall that the time axis is logarithmic, so constant star formation rate corresponds to a line of slope unity
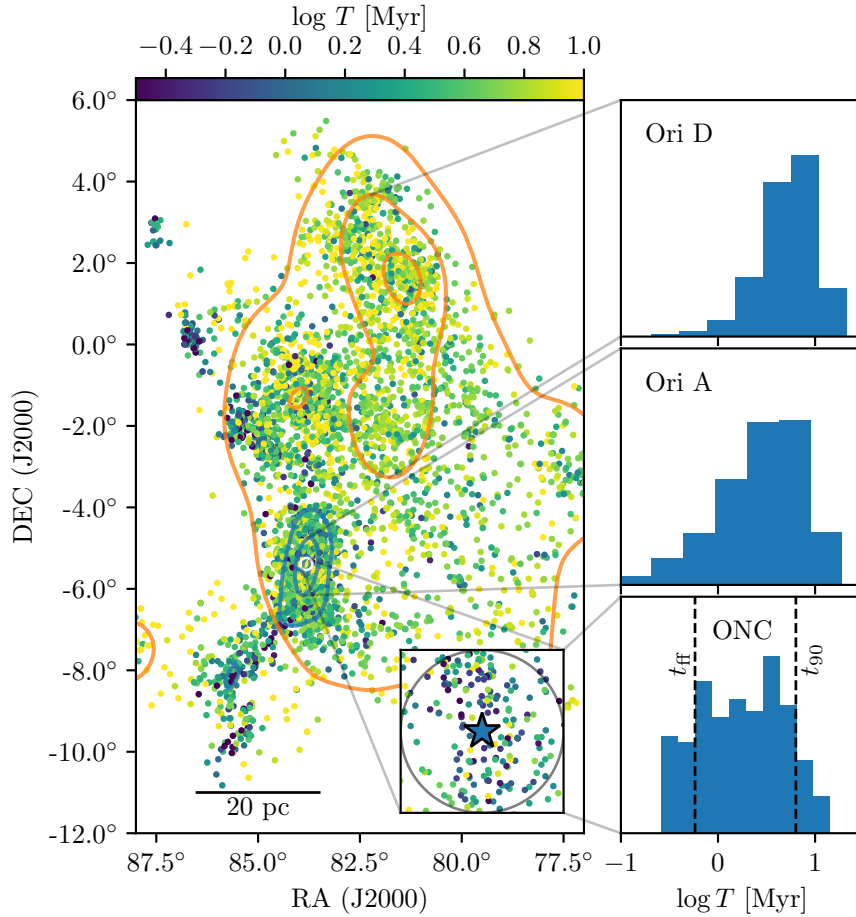
Figure 13

Distribution of stellar ages in the Orion star-forming complex. In the main panel, points are stars color-coded by age as estimated by Kounkel et al. (2018). The orange and blue contours, respectively, show the density of stars on the sky associated with the components Orion D and Orion A identified in Kounkel et al.'s phase-space analysis; contours are placed at densities of 10%, 50%, and 90% of maximum density. The inset shows a zoom-in on the ONC, defined here as a 1 pc-radius region centered on θ^1 C (shown by the star). The flanking histograms show the age distributions for Orion D, Orion A, and the ONC, as indicated. In the ONC panel, we also show t_{ff}, the free-fall time in the central pc of the ONC as estimated by Da Rio, Tan & Jaehnig (2014), and $t_{90} \approx 10t_{ff}$, the time over which 90% of the stellar population formed. The ages we show are those estimated based on spectroscopy (Kounkel et al.'s age$_{HR}$) where available, and based on color (Kounkel et al.'s age$_{CMD}$) elsewhere. However, in the ONC inset and histogram we show only spectroscopic ages, since the color-based ones are unreliable in regions of high extinction.

rather than zero.) Using the ages shown in Figure 13, we find that the time over which 90% of the stars formed is $\approx 10t_{ff}$, but roughly 50% of stars formed in the last $\approx 3t_{ff}$.

Third, mean age varies with position. The stars in the ONC are significantly younger than the mean

FALSE AGE SPREADS FOR ACCRETING YSOS?

Some authors have suggested that the dispersion in YSO luminosity at fixed effective temperature, usually interpreted as an age spread, might instead be due to variation in YSO accretion histories (e.g., Baraffe, Chabrier & Gallardo 2009; Baraffe, Vorobyov & Chabrier 2012). If some YSOs were to undergo exclusively "cold" accretion, meaning that accreting material radiates away all its entropy at the accretion shock, they would descend the Hayashi track more rapidly than YSOs with less efficient shocks, leading to a luminosity spread that could be mistaken for an age spread. However, the current consensus is that observed stellar age spreads are *not* due to this effect. Although both cold accretion and Kelvin-Helmholtz contraction drive stars down the Hayashi track, they do so at very different rates as a function of stellar mass. As a result, rather than generating a mass-independent age spread, cold accretion would make it appear that stars with effective temperatures $\gtrsim 3500$ K are systematically older than cooler stars (Hosokawa, Offner & Krumholz 2011). For these higher T_{eff} stars, cold accretion would also produce a small population that would be extremely luminous and thus appear young (Vorobyov et al. 2017; Jensen & Haugbølle 2018), rendering the overall apparent age distribution bimodal. Neither a bimodal inferred age distribution for hotter stars nor a systematic difference in inferred age as a function of effective temperature are observed, which puts strong limits on how much of the apparent age spread is due to cold accretion.

of the larger Orion A region, which is in turn younger than the still-larger Orion D region. Consequently, there is a gradient of increasing mean age and age spread as one moves from denser to less dense regions (Getman, Feigelson & Kuhn 2014; Getman et al. 2018). However, age does not increase as quickly as free-fall time, so while the ONC is ≈ 10 free-fall times old, the lower density regions around it in Orion A are only $\sim 1 - 3$ free-fall times old (Da Rio, Tan & Jaehnig 2014; Jaehnig, Da Rio & Tan 2015). Thus the outskirts of clusters are older than their centers in an absolute sense, but are dynamically younger. This statement likely applies to Orion D as well, but we cannot be sure because the gas has already been cleared, so the stellar density only provides an upper limit on the free-fall time.[13]

3.4.2. Proposed scenarios. For an isolated star-forming cloud, the natural evolutionary timescale is t_{ff}, and simulations of clouds that do not include an external environment (or effectively mock one up by using periodic boundary conditions) generally find that all gas is either converted to stars or expelled on such timescales (e.g., Krumholz, Klein & McKee 2012; Kim, Kim & Ostriker 2018; Grudić et al. 2018, though for an exception see Wang et al. 2010). Thus the observed spread of $\approx 10t_{\mathrm{ff}}$ in the ONC is inconsistent with expectations from isolated cloud models.[14] Observed star formation histories also present a second challenge: even with $t_{\mathrm{sf}} \approx 10t_{\mathrm{ff}}$ as in the ONC, since $\epsilon_{\mathrm{ff}} \approx 0.01$ (c.f. § 3.2) the ONC should reach a maximum conversion efficiency $\epsilon_* \approx 0.1$, probably too low to yield a gravitationally-bound cluster (c.f. § 3.5). It is possible that the ONC will in fact not remain bound (Kroupa, Aarseth & Hurley 2001), though recent *Gaia*

---

[13]This observation reinforces the problem with cloud matching identified in § 3.2: cloud matching would likely attribute the stars in Orion D to the present-day molecular gas cloud that is confined to the Orion A region. However, kinematic measurements demonstrate that the stars in Orion D did not form in the Orion A cloud, but from a separate, now-dispersed reservoir of molecular gas.

[14]An analogous problem occurs on the scales of entire GMCs, where simulations of isolated GMCs with properties similar to observed ones generally yield lifetimes of $\lesssim 5$ Myr (e.g., Grudić et al. 2018), while observationally-estimated GMC lifetimes are closer to 20-30 Myr (Dobbs et al. 2014).

observations suggest that it will (Kuhn et al. 2018; Kim et al. 2019). Moreover, the general point holds regardless of the fate of the ONC: the observed low median value of ϵ_{ff} means that bound clusters are likely to arise only from regions where ϵ_{ff} is in the upper $\sim 20\%$ of the distribution (plausible, since only $\sim 10\%$ of star formation results in bound clusters), where the free-fall time is short enough for the star formation history to extend over > 10 free-fall times (Kruijssen 2012), or some combination of the two.

One possible explanation for the ONC's star formation history is that it did form in a single free-fall time, but in an extended region of lower density than that the stars currently occupy. This region was in a free-fall collapse that halted when massive stars dispersed the gas. In this scenario, $t_{\mathrm{sf}} \gg t_{\mathrm{ff}}$ not because star formation was extended in time, but because t_{ff} decreased after the stars formed (e.g., Zamora-Avilés & Vázquez-Semadeni 2014; Kuznetsova, Hartmann & Ballesteros-Paredes 2015, 2018; Vázquez-Semadeni, González-Samaniego & Colín 2017). This would naturally give rise to an age gradient such as that observed, because stars that formed earlier in the process would retain a higher proportion of their kinetic energy, and thus end up further from the cluster center in the final configuration (Getman et al. 2018).

However, this scenario encounters serious observational difficulties. First, it would produce a distinct kinematic signature that is not observed. During collapse stellar motions should be radially inward, and once collapse is complete and gas has been expelled, the $\sim 90\%$ of the stars that do not remain bound (since $\Gamma \sim 10\%$; § 2.4) should be moving radially outward from the dense center; in either case stellar velocities should point radially toward or away from the densest point (in the example of Figure 13, the ONC). However, *Gaia* shows that this is not the case in Orion or in other complexes (Ward & Kruijssen 2018; Kuhn et al. 2018; Kounkel et al. 2018). Note that this radial expansion is not seen in the simulations of Kuznetsova, Hartmann & Ballesteros-Paredes (2015, 2018), but this is because the simulations do not include any form of feedback, and thus nearly all the stars they produce remain bound. If the stars in their final configuration were to become unbound somehow, the velocities would become largely radial as soon as the stars expanded significantly compared to their original volume. A second problem is that this scenario would require even more extreme efficiencies, $\epsilon_{\mathrm{ff}} \gtrsim 0.3$, to produce bound systems; excluding the discrepant cloud matching results, observed ϵ_{ff} values do not go this high (c.f. § 3.2). A third problem is that this scenario requires star clusters to pass through a dense phase wherein the mass is still largely in the form of gas, so that for any given star cluster, there should be a comparably dense gas cloud. However, surveys have consistently failed to find gas clouds as dense as the densest YMCs (Ginsburg et al. 2012; Longmore et al. 2014; Walker et al. 2016; Urquhart et al. 2018). We illustrate this with a current compilation of gas clouds and clusters in Figure 14. Notice that, for the densest clusters in the disk of the Galaxy (though interestingly not for the Galactic Center), there are no clouds dense and massive enough to be their progenitors if the entire cloud had to be assembled at once. Given these problems, it is unlikely that the extended formation times are illusory.

How then are the long formation times to be explained? One possibility is that existing simulations do not yet have enough physics or enough resolution to produce realistic lifetimes. There would be some precedent for this: for decades simulations struggled to reproduce the small observed values of ϵ_{ff}, but, as noted in § 3.2, simulations that include multiple feedback processes, initial turbulence, and magnetic fields are now approaching this goal. It is possible that the same will happen with the duration of star formation. More likely, however, the problem with current simulations is lack of the larger scale-environment that provides an ongoing mass supply. Observed protoclusters are generally located at the confluence of gaseous filaments, suggesting a "conveyor belt" picture in which gas accretion onto a protocluster and star formation occur simultaneously (Longmore et al. 2014; Motte, Bontemps & Louvet 2018; see Supplementary Materials). Inflow supplies energy to maintain turbulence and prevent contraction (Klessen & Hennebelle 2010; Goldbaum et al. 2011; Matzner & Jumper 2015; Lee & Hennebelle 2016a,b), which together with local feedback processes keeps ϵ_{ff} low. This state persists as long as the accretion lasts, which can be many times
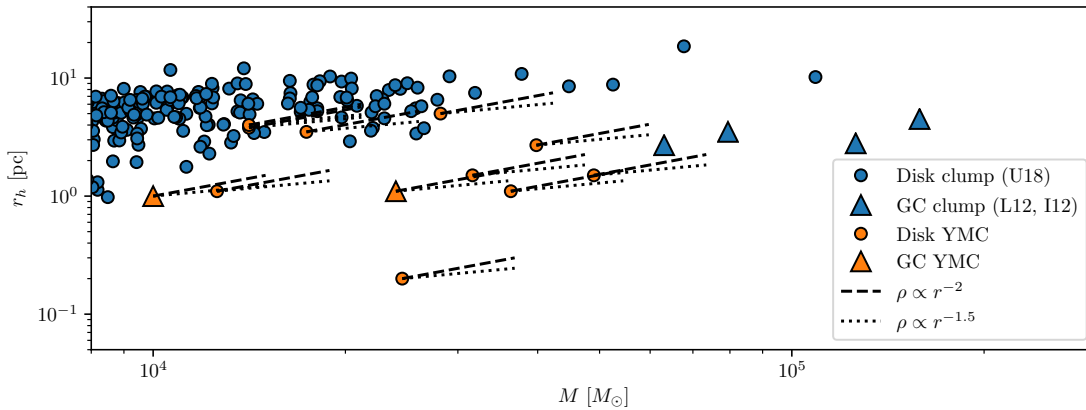
Figure 14

Comparison of masses and radii of young massive clusters (YMCs) in the disk of the Milky Way and near the Galactic Center, as indicated, and for clumps of molecular gas. The YMCs plotted are those listed in Table 3. For gas clouds in the Galactic disk, we show the sample of Urquhart et al. (2018, U18) from ATLASGAL, which is complete except near the Galactic Center. We supplement this with the Galactic Center clouds reported by Longmore et al. (2012, L12) and Immer et al. (2012, I12). The dashed and dot-dashed lines marked for each cluster show the mass as a function of radius for clouds with density as a function of radius $\rho \propto r^{-2}$ and $r^{-1.5}$, as indicated. These lines indicate the loci below which progenitors clouds for these clusters would have to lie if their densities varied as the indicated power-law, i.e., a gas cloud with density versus radius $\rho \propto r^{-2}$ would only be dense enough to form the indicated star cluster if it were to lie below the dashed line emerging from that star cluster. For several disk clusters, there are no gas clouds that satisfy this condition.

the local value of $t_{\rm ff}$ because the accretion flow is dictated by processes occurring on much larger scales with much longer dynamical times. Because $\epsilon_{\rm ff}$ is small and gas consumption slow, the star-forming gas mass grows with time, leading to an accelerating star formation rate. Once accretion ceases, and the turbulent energy supply and confining ram pressure of the accretion are removed, any remaining gas is converted to stars or expelled by feedback in a time of order $t_{\rm ff}$, as happens in the isolated cloud simulations. The final star formation efficiency is ultimately be dictated by the efficiency of feedback at ejecting gas, and if the ratio of mass ejected by feedback to mass converted into stars is $\lesssim 1$, then the final value of ϵ_* is large even if $\epsilon_{\rm ff}$ is always fairly small.

This scenario is consistent with observed values of $\epsilon_{\rm ff}$, naturally produces an accelerating rate of star formation, and also yields a cluster mass-size relation in good agreement with observations (Matzner & Jumper 2015; Lee & Hennebelle 2016b). Because the final mass that goes into a cluster was never assembled in a single cloud all at once, there is no problem with finding clouds dense enough to be the progenitors of the densest clusters. It is not clear if this mechanism is capable of producing the core-halo age gradient observed by Getman, Feigelson & Kuhn (2014) and Getman et al. (2018), and visible in Figure 13 – for the ONC this gradient is ~ 0.5 Myr pc^{-1}. Such a gradient could arise as a result of earlier-forming stars having longer to undergo dynamical heating in the core of the forming cluster, but this possibility has not been examined quantitatively. A second possible challenge for this picture is that preliminary work suggests that it leads to clusters with significant rotational support Lee & Hennebelle (2016a,b). While some clusters do rotate (e.g. Davies et al. 2011; Hénault-Brunet et al. 2012; Kamann et al. 2018), others do not (Kuhn et al. 2018), and it is not clear if the predicted rotation rate distribution is consistent with observations. However, current theoretical predictions for the rotation rate distribution do not yet take into account the

possibility of braking by dynamically-significant magnetic fields.

3.5. Emergence of the gas-free cluster population

3.5.1. The initial cluster mass function. The initial cluster mass function (ICMF), defined as the CMF immediately after gas removal, is the starting point for understanding the evolution of clusters. The ICMF is less well defined than its stellar counterpart because the stellar content of a cluster is constantly evolving, particularly for unbound clusters. Adams & Myers (2001) showed that a cluster must contain $\gtrsim 100$ stars in order for the dynamical time to exceed the formation time, which they took to be $\sim 1 - 2$ Myr. Indeed, the ICMF for embedded clusters in the Solar neighborhood turns down below $N \sim 100$ stars, corresponding to a cluster mass $\sim 50\,M_\odot$ (Lada & Lada 2003). As discussed in § 2.2, the CMF for observed clusters usually can be fit with a Schechter function with a power-law index $\alpha_M \approx -2$ and a truncation mass M_c that is related to the environment. Elmegreen (2006) has shown that a slope $\alpha_M = -2$ naturally accounts for the fact that the stellar IMF averaged over a galaxy is very close to the average IMF of individual star clusters.

3.5.1.1. The clump mass function. Theoretically, the ICMF is the product of the clump mass function (§ 3.1.2) and the mass-dependent star formation efficiency $\epsilon_*(M)$. This is approximate: in a turbulent medium the clump mass evolves with time, and we have seen in § 3.4 that bound clusters form over multiple dynamical times with continuous accretion, so that the star-forming gas is never assembled into a single monolithic clump. Nonetheless, clumps still provide a useful approximation for modeling the complex process of cluster formation, just as protostellar cores provide a useful basis for modeling the stellar IMF (e.g., Hennebelle & Chabrier 2008). An important factor determining the clump mass function, $dN_{\rm clump}/dM \propto M^{\alpha_{M,{\rm clump}}}$, is that molecular clouds are supersonically turbulent and appear to be self-similar from the outer scale of the turbulence (generally encompassing the entire cloud) to the sonic scale (~ 0.1 pc in Galactic GMCs). Fleck (1996) was the first to argue that $\alpha_{M,{\rm clump}} = -2$ for a self-similar medium in which clumps form hierarchically. Elmegreen & Falgarone (1996) came to a similar conclusion by using the observed fractal dimension of molecular clouds, and for the first time suggested that this would lead to a similar slope for the ICMF. Guszejnov, Hopkins & Grudić (2018) have elaborated the argument for $\alpha_{M,{\rm clump}} = -2$, which we paraphrase: assume that the clouds form hierarchically, either by fragmentation or by mergers, and that each level of the hierarchy, n, contains N_n clumps that have a range of masses $\Delta M_n \propto M_n$. Then the total mass at level n is

$$M_{\rm tot}(M_n) = M_n N_n = M_n \left(\frac{N_n}{\Delta M_n} \right) \left(\frac{\Delta M_n}{M_n} \right) M_n \propto M_n^2 \frac{dN}{dM}, \tag{14}$$

where the last step follows for large n. If the hierarchy is populated throughout, so that the mass at each level of the hierarchy, $M_{\rm tot}(M_n)$, is about constant, then $\alpha_{M,{\rm clump}} \approx -2$. This argument is clearly approximate; for example, GMCs have most of their mass at high masses in the Milky Way and at low masses in some other galaxies (§ 3.1.2). In a dynamical system such as a GMC, the clumps at different levels of the hierarchy can also have different lifetimes, thereby changing the observed PDF; for example, if the clumps have mass-independent surface densities and survive for a time proportional to the free-fall time, the observed value of $\alpha_{M,{\rm clump}}$ is larger than the intrinsic value by $1/4$ (Fall, Krumholz & Matzner 2010). This theory of the clump mass function is thus not very precise, but it is broadly consistent with observation.

3.5.1.2. From clumps to clusters. As noted above, the ICMF is related to the mass distribution of the natal molecular clumps by the star formation efficiency (SFE), $\epsilon_*(M)$, so that $dN(M_*) \propto \epsilon_*(M)dN_{\rm clump}(M)$ approximately, where it is possible that more than one cluster forms in a cloud. Photoionization and direct

radiation pressure are the most broadly effective feedback mechanisms (§ 3.3). Photoionization removes a mass $\Delta M_{\rm ion} \propto S^{4/7} t^{9/7}/\rho^{1/7}$ over a time t, where S is the rate of production of ionizing photons (Whitworth 1979). If this is the dominant feedback mechanism, star formation will continue until $\Delta M_{\rm ion} \sim M$. In terms of ϵ_*, this will occur at the star formation time $t_{\rm sf} = \epsilon_* t_{\rm ff}/\epsilon_{\rm ff}$, so that

$$\epsilon_* \propto \epsilon_{\rm ff}^{9/13} \Sigma^{33/52} M^{1/52}. \tag{15}$$

Bound clumps ($\alpha_{\rm vir} \sim 1$) have surface densities that depend only on the turbulence parameter, $C = \sigma/R^{1/2}$ (Equation 2), which is observed to be approximately constant in a galaxy (e.g., Sun et al. 2018). For constant $\epsilon_{\rm ff}$ (§ 3.2), this implies ϵ_* is nearly independent of M, so the ICMF has the same slope as the clump mass function. If on the other hand stars form in a fixed number of free fall times so that $\epsilon_* \propto \epsilon_{\rm ff}$, then $\epsilon_* \propto \Sigma^{33/16} M^{1/16}$, very close the results of Kim, Kim & Ostriker (2018) for small ϵ_*.

Direct radiation pressure injects momentum at a rate proportional to the stellar mass. Stars will form until the injected momentum $\propto \epsilon_* M t_{\rm sf}$ reaches $M v_{\rm esc}$, so that

$$\epsilon_* \propto (\epsilon_{\rm ff} \Sigma)^{1/2}, \tag{16}$$

with no explicit dependence on M. If $t_{\rm sf} \propto t_{\rm ff}$, so that $\epsilon_* \propto \epsilon_{\rm ff}$, then $\epsilon_* \propto \Sigma$, as found by Fall, Krumholz & Matzner (2010). In this case, or if $\epsilon_{\rm ff}$ is about constant, the SFE is independent of mass for bound clumps, which have similar surface densities in a given galaxy as noted above. As a result, the slope of the ICMF is the same as that for the clump mass function, $\alpha_{M,\,{\rm clump}} \approx \alpha_M \approx -2$.

3.5.2. Bound clusters. With our broad definition of cluster (§ 1.2.2), essentially all stars are born in clusters (c.f. Lada & Lada 2003). However, most clusters are unbound by gas dispersal and thus disperse rapidly, so that $\Gamma \lesssim 0.1$ for $T \gtrsim 10$ Myr. In this section we investigate the origin of the low value of Γ after gas dispersal. As discussed in § 3.1.1, most GMCs appear to be bound (i.e., $\alpha_{\rm vir} \lesssim 2$), as are the dense clumps within them (Urquhart et al. 2018), so most stars are born in gravitationally bound gas. Even in GMCs that are moderately unbound, $\alpha_{\rm vir} \approx 5$, efficient star formation results in most stars being confined to gravitationally bound structures (Clark et al. 2005). Thus the low value of Γ must be related to the star formation efficiency, ϵ_*, and to how the gas not converted to stars is ejected.

The first author to estimate Γ considered a spherical cloud in virial equilibrium that turns a fraction ϵ_* of its mass into stars and ejects the rest (Hills 1980). If gas removal occurs over many dynamical times, the process is adiabatic, the cluster remains bound for any ϵ_*, and its final and initial radii are related by $R_f/R_0 = 1/\epsilon_*$. If gas removal is rapid, ejection of more than half the mass ($\epsilon_* < 0.5$) unbinds the stars; for $\epsilon_* > 0.5$ the radius expands, $R_f/R_0 = \epsilon_*/(2\epsilon_* - 1)$. Even for $\epsilon_* < 0.5$, however, the low-velocity part of the cluster can remain bound (Lada, Margulis & Dearborn 1984). In simulations, Kroupa, Aarseth & Hurley (2001) find bound remnants remain for ϵ_* as small as 0.3.

More realistic starting conditions can modify these conclusions. If the stars are initially subvirial, with $Q_0 = 2T_*/|W_*| < 1$, where T_* and W_* are the initial kinetic and potential energies of the stars, respectively, then one can show that the minimum ϵ_* for the entire star cluster to be bound is $\epsilon_{*,\min} = Q_0/2$. If a cluster has substructure, then the additional binding energy can help it survive (Allison et al. 2009); whether the cluster remains bound depends unpredictably on the exact spatial distributions of gas and stars and on the timing of gas removal (Smith et al. 2011, 2013). The minimum value of ϵ_* for a bound cluster is also reduced if the stars are more concentrated than the gas, so that the local value of ϵ_* increases inward. Adams (2000) showed that part of a cluster embedded in a gas distribution approximating an isothermal sphere could survive sudden gas expulsion down to very low values of ϵ_*. A cluster that forms in a centrally concentrated gas cloud with a constant value of $\epsilon_{\rm ff}$ has a star formation rate $\dot{\rho}_* \propto \rho^{3/2}$ (Parmentier & Pfalzner 2013); in a particular implementation of this model, Shukirgaliyev et al. (2017) found a minimum

global star formation efficiency $\epsilon_{*,\mathrm{min}} = 0.15$ for some stars to remain bound. Analysis of a simulation of the formation of a cluster led Kruijssen et al. (2012a) to conclude that most cluster stars form in gas-poor subclusters that have locally high ϵ_* and are therefore relatively immune to gas expulsion; however, this work did not include feedback or magnetic fields, and thus its ϵ_* values are unrealistically high.

Real star-forming regions are hierarchically-structured, containing both dense parts for which mass removal is slow compared to the local dynamical time, and diffuse parts for which it is fast; we see this directly in the Orion star forming complex (§ 3.4). Elmegreen (2006, 2008) was the first to point out that this configuration naturally gives rise to a gravitationally-bound central region and an unbound periphery. Kruijssen (2012) provided a quantitative model for Γ in such an environment, based on three main premises: first, the density PDF is lognormal, as expected in supersonically-turbulent flows; second, $\epsilon_{\mathrm{ff}} \approx 0.01$ independent of density, as observations suggestion (§ 3.2); third, gas at all densities is removed rapidly after some fixed timescale t_{sf} set by stellar feedback. Given these ingredients, he calculated a density-dependent star formation efficiency $\epsilon_*(\rho) = \epsilon_{\mathrm{ff}}[t_{\mathrm{sf}}/t_{\mathrm{ff}}(\rho)]$, and integrated over the density PDF to derive the fraction of mass for which $\epsilon_* \gtrsim 0.5$, which he associated with the bound fraction Γ. A generic prediction of this model is that Γ increases with mean ISM density or pressure, although such a prediction is not unique to this model, since many of the feedback mechanisms discussed in § 3.3 become less effective at higher gas surface density.

In this view, the low value of Γ is not so much "infant mortality" as "infant weight loss": stars form in hierarchical structures where only a small fraction remains bound after gas dispersal, so rather than $\Gamma = 0.1$ meaning that 90% of clusters disperse and 10% remain bound, it means that 90% of the stellar mass is in the unbound part of the structure that escapes immediately after gas dispersal. Of course in reality both mortality and weight loss may operate: hierarchical structures may lose much of their stellar mass in low-density regions, and then even the denser regions that survive may disperse soon thereafter if they are insufficiently concentrated to survive stellar evolution and tidal shocking (§ 4).

Although the case in which stars form in a pre-existing density structure (either uniform or hierarchically-structured) has garnered the most attention in the literature, we have argued that it is more likely star formation and cloud assembly are contemporaneous. One can generalize the star formation model in § 3.2 to allow for gas accretion while the stars form (see "Conveyor-Belt Model for Cluster Formation" in the Supplementary Materials). However, provided the time over which the cloud accretes is $\lesssim t_{\mathrm{sf}}$ (Equation 10) the results are similar to those above. If the accretion stops due to feedback, then it is likely that the subsequent mass loss would be fast, since the same processes stopping the accretion would eject the gas in the clump. On the other hand, if the accretion stops because the supply of new gas runs out, then stars would continue to form and the mass loss would be slow.

One final consideration is that clusters form in tidal gravitational fields (§ 4.1), which sets a minimum density for the bound part of the cluster, ρ_{ti}. After mass loss, the final mean density of a cluster with initial radius R_0 and density ρ_0 is (Mathieu 1983)

$$\frac{\rho_{*f}}{\rho_0} = f_b \epsilon_* \left(\frac{R_0}{R_f}\right)^3 \rightarrow \begin{cases} \dfrac{(2\epsilon_* - 1)^3}{\epsilon_*^2} & \text{fast mass loss,} \\ \epsilon_*^4 & \text{slow mass loss,} \end{cases} \tag{17}$$

where f_b is the fraction of the stellar mass that is bound, and we have assumed $f_b = Q_0 = 1$ in the expressions for the limiting cases.[15] Clusters are generally centrally concentrated, and stars in the outer parts of the cluster in which the final stellar density, ρ_{*f}, is less than the tidal limit will be lost. For a simple spherically symmetric model in which (1) the gas density varies as r^{-k} with $k \simeq 2$ (Schneider et al.

---

[15]The expression for the fast case is valid only for $\epsilon_* \gtrsim 0.6$ (Lada, Margulis & Dearborn 1984).

2015), (2) $\epsilon_* \propto 1/t_{\mathrm{ff}}$ (as in the Kruijssen 2012 model), so $\epsilon_* \propto \rho^{1/2} \propto 1/r$, and (3) ϵ_* near the center approaches unity (see "Cluster-Forming Clouds and Clumps" in the Supplementary Materials), the bound fraction is large, $f_b \gtrsim 0.5$, for values of the cloud-averaged $\epsilon_* \gtrsim 0.01$; in this model, the mean value of ϵ_* for the stars that remain bound is $\epsilon_{*b} \gtrsim 0.3$.[16]

Altogether, theoretical estimates of the bound fraction are the product of several uncertain factors. First, after the accretion of gas onto the cluster-forming clump has ceased, what fraction of the clumps are bound? Second, what fraction of the stars remain bound after the residual gas is ejected? Observations suggest that star formation is coincident with clump formation, at least in the early stages, but it is not clear whether this process is halted by feedback, leading to rapid mass loss and requiring a relatively high star formation efficiency, or by a lack of gas, which would be consistent with a lower efficiency. Tides due to the local environment can also strip stars from the outer regions of the cluster when mass is lost, a point to which we return in § 4.3.1.

3.5.3. Numerical results.
A few authors have attempted to simulate the full range of processes – formation of a mass spectrum of clumps, conversion of gas into stars, and gas removal by feedback – that determine the ICMF and Γ. This is very challenging numerically, since capturing the formation of a statistically-meaningful sample of clusters requires simulating either an entire galaxy or a substantial portion thereof. Modern isolated galaxy or cosmological zoom-in simulations generally reach spatial resolutions no better than ~ 1 pc, and mass resolutions of $\sim 10^2 - 10^3 \ M_\odot$. Comparing these figures to the star cluster properties shown in Figure 9, even for GCs this corresponds to $\lesssim 10^3$ mass resolution elements per cluster, and $\lesssim 3-10$ spatial resolution elements per half-mass radius. Thus simulations must rely on parameterized subgrid models to handle both star formation and feedback.

Perhaps not surprisingly, this situation yields little consensus. Renaud, Bournaud & Duc (2015) simulate an Antennae-like galaxy merger, and find that Γ is a few percent pre-merger, but rises to $\approx 10\%$ during the main burst of star formation. In contrast, Li, Gnedin & Gnedin (2018) find Γ values from $\approx 1\%$ to $\approx 50\%$ in their cosmological simulations that run to $z \approx 1.5$. Renaud, Bournaud & Duc (2015) and Maji et al. (2017) find lognormal ICMFs in simulations of both mergers and quiescent spiral galaxies (contrary to observations), while Dobbs et al. (2017), Li et al. (2017), and Li, Gnedin & Gnedin (2018) find ICMFs that are well-described by power-laws or Schechter functions with $\alpha_M \approx -2$. Li, Gnedin & Gnedin (2018) test a variety of star formation and feedback prescriptions, all tuned to reproduce the galaxy-scale Kennicutt relation and various other constraints, and find that both Γ and the ICMF are extremely sensitive to the choice of subgrid model.

We conclude that present galaxy simulations have little predictive power when it comes to star clusters; instead, a more productive approach would be to use the observed properties of star clusters as an additional constraint to calibrate the subgrid models. There do, however, appear to be two qualitative results that persist across recipes. First, all simulations find that Γ increases at higher surface densities of star formation, which occur in regions of higher pressure. Second, all authors find that the ICMF extends to higher masses during mergers. This is also consistent with the results of a number of other cosmological simulations showing that early, gas-rich galaxies are capable of producing clusters with initial masses $\gtrsim 10^6 \ M_\odot$ (Kimm et al. 2016; Ricotti, Parry & Gnedin 2016; Kim et al. 2018).

---

[16]There are several distinct bound fractions: Γ is the fraction of all stars that are bound, f_b is the fraction of stars that are bound in a particular cloud, and ϵ_{*b} is the fraction of gas that went into stars in the bound part of a cluster.

3.6. Elemental abundance distributions

There has been surprisingly little theoretical work on the physical origins of abundance homogeneity in star clusters. Murray & Lin (1990) observed that narrow giant branches in GC CMDs implied they were highly uniform in [Fe/H], and suggested that turbulent diffusion could explain why clouds are highly uniform prior to the onset of star formation. This process would mix out abundance inhomogeneities at the scale of a cloud over a cloud crossing time, and smaller-scale inhomogeneities even more quickly. Feng & Krumholz (2014) provide numerical support for this idea, finding in a simulation of star formation in a colliding flow of warm neutral gas that the mixing process washes out homogeneities in the initial colliding streams by a factor of 5 or more. This may be telling us that the small inhomogeneities detected in a few open clusters (e.g. M67; Liu et al. 2016b) reflect metal anomalies (predicted to be ~ 0.1 dex – Krumholz & Ting 2018) in the birth cloud (§ 2.6). Feng & Krumholz also find that even unbound clusters can also be homogeneous, a prediction that has yet to be tested by observations. Armillotta, Krumholz & Fujimoto (2018) revisit the Feng & Krumholz (2014) simulations using more realistic initial conditions, a cloud extracted from a galaxy simulation, and zoom in to follow it collapse into individual stars at a resolution of $\approx 10^{-3}$ pc. They find that the abundance scatter between formed stars decreases on shorter length scales, indicating that turbulent mixing becomes more efficient on smaller and smaller scales ($\lesssim 5$ pc), in agreement with the discussion above. Interestingly, they also find that some stellar clusters are separable from others within the collapsing clouds based on abundances, but not all.

A few caveats are in order. First, the simulations of Feng & Krumholz (2014) and Armillotta, Krumholz & Fujimoto (2018) assume that metals are well-coupled to the gas, but this need not be the case for metals in the form of dust grains. Padoan et al. (2006) first raised the prospect that small-scale turbulence could generate low-level variations in the dust to gas ratio. More recently, Hopkins & Lee (2016) and Lee, Hopkins & Squire (2017) find far more dramatic variations in their simulations and even suggested that some stars could form in "totally metal" regions (Hopkins 2014). But this work has been refuted by Tricco, Price & Laibe (2017) who find no evidence for significant variations; the use of tracer particles in supersonic flows lead to numerical artifacts that exaggerate density contrasts. Second, the cluster formation simulations to date have not considered the effects of magnetic fields, which Sur, Pan & Scannapieco (2014) suggest can suppress the mixing over a wide range in Mach number. Here, the same random stretching and folding that is responsible for turbulent diffusion also amplifies the magnetic energy density by the dynamo mechanism, which in turn can suppress the mixing on small scales.

A third caveat is that, while galaxy-scale simulations are starting to consider injection of new metals by stellar evolution (e.g., Emerick, Bryan & Mac Low 2019), simulations that resolve cluster formation have yet to do so. Bland-Hawthorn, Krumholz & Freeman (2010) point out that, even if turbulence homogenizes a cloud, it will only remain homogeneous if star formation ceases before the first SN explodes, since even a single SN will produce measurable abundance variations. For a gas cloud with mass M, surface density Σ, and virial ratio α_{vir}, the crossing time is $t_{\mathrm{cr}} \approx (\alpha_{\mathrm{vir}} G)^{-1/2} M^{1/4} \Sigma^{-3/4}$ (c.f. § 3.1.1). If we adopt $\alpha_{\mathrm{vir}} = 1.5$, and the cluster forms over $\tau_{\mathrm{sf}} = t_{\mathrm{sf}}/t_{\mathrm{cr}}$ crossing times, then

$$t_{\mathrm{sf}} \approx 3.0 \left(\frac{\tau_{\mathrm{sf}}}{4}\right) \left(\frac{\epsilon_*}{0.2}\right)^{-1/4} \left(\frac{M_*}{10^4\,M_\odot}\right)^{1/4} \left(\frac{\Sigma}{0.3\ \mathrm{g\ cm^{-2}}}\right)^{-3/4}\ \mathrm{Myr}, \tag{18}$$

where $\epsilon_* = M_*/M$ is the star formation efficiency. Thus $t_{\mathrm{sf}} < t_{\mathrm{SN}}$, which is conservative given that most SN progenitors have longer lifetimes, holds for OCs ($\Sigma \approx 0.3\,M_\odot$ pc^{-2}) up to $10^5 M_\odot$, and for GCs ($\Sigma \approx 3\,M_\odot$ pc^{-2}) up to $10^7 M_\odot$. This last estimate is interesting in light of the fact that globulars can show high levels of Fe homogeneity; as we have seen, for other elements, the story is more complicated. How there can be differential scatter across elements is not understood, unless it reflects the state in the collapsing cloud (see below). In Figure 9, there is evidently a lot of scatter but our point is to emphasize the difference between

OCs and denser GCs.

Both this calculation and the simulations of Armillotta, Krumholz & Fujimoto (2018) raise the question of how to apply estimates of chemical homogeneity or inhomogeneity in the context of hierarchical star-forming systems that do not have clear edges. This topic is currently under investigation using a combination of *Gaia* data and ground-based, million-star surveys at high spectroscopic resolution (e.g. GALAH; De Silva et al. 2015). The astrometric information can be used to identify every aggregate within dozens of nearby stellar hierarchies (Gouliermis 2018), while the spectra provide abundances for all stars cooler than ≈ 7000 K. This work will provide the first clues for how to map stars between abundance and physical space, i.e., the possibility of determining how closely together in space two stars formed based on the level of similarity or difference in their elemental composition, with major implications for the future of chemical tagging in Galactic archaeology (Freeman & Bland-Hawthorn 2002; Bland-Hawthorn & Freeman 2004). It is also not clear how to apply these models in the context of multiple generations and detectable abundance anomalies in globular clusters (§ 2.6).

4. LIFE AND DEATH

Since the slope of the CAF $\alpha_T < 0$ even for ages older than ≈ 10 Myr (§ 2.3), cluster disruption cannot depend solely on gas expulsion immediately after cluster formation. Instead, there must be processes that destroy star clusters during the gas-free phase of their evolution. It is to these processes that we now turn. We begin with a brief review of star cluster structure and tidal fields in § 4.1, then examine disruption processes driven primarily by mechanisms internal to the cluster (§ 4.2) and those driven primarily by the external environment (§ 4.3), before concluding with attempts to synthesize both into a coherent explanation for the demographics of star clusters § 4.4. Since our focus is on the demographics and evolution of the star cluster population as a whole, we will not discuss processes that alter the internal structure of star clusters while leaving their bulk properties unchanged, e.g., dynamical alteration of the binary fraction or stellar collisions. We refer readers to Portegies Zwart, McMillan & Gieles (2010) and Renaud (2018) for a discussion of these processes.

4.1. Star cluster structure and tides

A crossing time after gas is removed, star clusters that remain bound relax to a nearly-spherical, virialized state. From this point up to ages of several hundred Myr, their density profiles are well described by the Elson, Fall & Freeman (1987, EFF hereafter) distribution, which has surface density as a function of radius $\Sigma(r) = \Sigma_0 \left(1 + r^2/a^2\right)^{-\gamma/2}$. Here Σ_0 is the central surface density, a defines the radial scale, and γ describes the fall off in surface density in the outer, power-law region; observed values of γ are generally in the range $\approx 2.2 - 3.2$ (Elson, Fall & Freeman 1987; Mackey & Gilmore 2003a,b). The core radius for an EFF profile, defined to match that of a King profile (see below) is $r_c \approx a\sqrt{2^{2/\gamma} - 1}$.

The EFF profile is not in energetic equilibrium, and thus clusters observed to have this distribution cannot have reached either internal relaxation or equilibrium with the tidal field of their host galaxy. At ages beyond a few hundred Myr, clusters relax and become well-fit by King (1962, 1966) models where stellar energies are distributed as a "lowered Maxwellian" of the form $f(E) \propto (2\pi\sigma^2)^{-3/2} \left[\exp\left(-E/\sigma^2\right) - 1\right]$ for some constant velocity dispersion parameter σ. Physically, the -1 in square brackets reflects the fact that the velocity distribution cannot be a pure Maxwellian, because the cluster has a finite escape speed, and thus any star that is too far out on the tail of the distribution will leave. Combining this assumed energy distribution with the Poisson equation for self-gravity yields an equation for the radial density distribution and potential. Solutions to this equation form a one-parameter family, with the parameter conventionally specified by setting the dimensionless potential well depth at the cluster center $W_0 = (v_{\mathrm{esc}}/\sigma)^2$, where
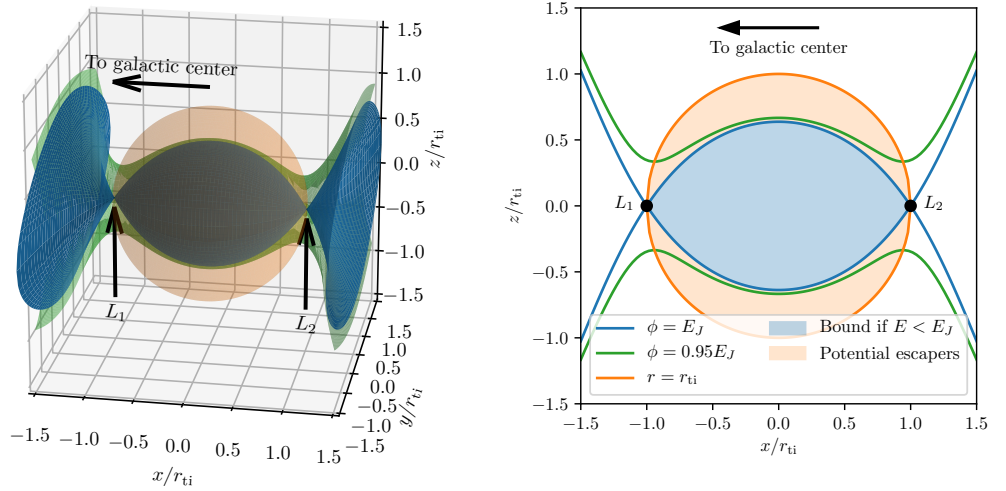
Figure 15

Isosurfaces of the potential $\phi(x, y, z)$ (Equation 19) for a point-like star cluster embedded in a galaxy, seen in 3d (*left*) and in a cut through the xz plane (*right*). The cluster center is located at the origin, while the galaxy center is located at $(-R_g, 0, 0)$ for some distance R_g much larger than the physical size of the cluster. In the example shown, $\lambda_y = 0$ and $\lambda_z/\lambda_x = -1/3$. Blue shows the Jacobi surface, defined by $\phi = E_J$, while green shows a potential $\phi = 0.95 E_J$, i.e., with energy 5% greater than the Jacobi energy. Orange shows a sphere of radius $r = r_{\rm ti}$ (Equation 20). The Lagrange points L_1 and L_2 are as indicated. Stars inside the Jacobi surface, shaded blue in the right panel, are bound cluster members if their total energy $E < E_J$. Stars outside the Jacobi surface but at $r < r_{\rm ti}$ (shaded orange), or those inside the Jacobi surface with $E \geq E_J$, are potentially unbound from the cluster, and can escape through the "windows" around L_1 and L_2.

$v_{\rm esc}$ is the escape speed from the cluster center. Most observed OCs and GCs are reasonably well-fit by $W_0 \approx 3 - 7$, though a small number of core-collapsed globulars and some of the most concentrated YMCs have $W_0 \gtrsim 10$.

For a King model, one can define the core radius $r_c = \sqrt{9\sigma^2/4\pi G \rho_0}$, where ρ_0 is the central density. The core radius is distinct from the half-mass radius r_h (§ 2.5), and r_h/r_c is a function of W_0, but for $W_0 \approx 3 - 7$, r_h/r_c is within a factor of a few of unity. A second useful quantity is the truncation radius $r_{\rm tr}$, the radius at which the density falls to zero. This can be combined with the core radius to yield the concentration $c = \log(r_{\rm tr}/r_c)$. This is a monotonically increasing function of W_0, so clusters with deeper potential wells are more centrally concentrated. For this reason, clusters are often parametrized in terms of values of c instead of values of W_0. The observed range $W_0 \approx 3 - 7$ corresponds to $c \approx 0.6 - 1.5$. Note that the EFF model does not have a truncation radius.

King models in principle apply to an isolated star cluster, but it is common to assume that the truncation at $r_{\rm tr}$ is due to the tidal field of the cluster's host galaxy.[17] One can derive an effective potential in the frame

---

[17]We emphasize that this assumption is not physically required. While a cluster in equilibrium cannot be larger than the size imposed by galactic tidal truncation, it could be smaller, e.g., if the cluster were truncated by gas removal unbinding its outer regions (c.f. § 3.5). Conversely, clusters can remain out of equilibrium and larger than their tidal radii for hundreds of Myr, which is why EFF profiles usually provide better fits than King profiles for

co-moving with a cluster orbiting with a galaxy by Taylor-expanding the galactic potential and adding this to the centrifugal acceleration in the rotating reference frame. If one approximates the cluster as point-like, the resulting total potential is (e.g., Renaud, Gieles & Boily 2011)

$$\phi(x, y, z) = -\frac{GM}{\sqrt{x^2 + y^2 + z^2}} - \frac{\lambda_x}{2}\left(x^2 + \frac{\lambda_y}{\lambda_x}y^2 + \frac{\lambda_z}{\lambda_x}z^2\right),$$ (19)

where M is the cluster mass, λ_x, λ_y, and λ_z are the three eigenvalues of the tidal plus centrifugal tensor,[18] and the coordinate system is oriented so that the center of the galaxy lies at on the x axis at $x < 0$, and the tidal plus centrifugal tensor has zero off-diagonal components. For a star cluster on a circular orbit at the midplane of an azimuthally-symmetric galactic potential ϕ_G, adopting a coordinate system where the y direction is spinward, the z direction is out of the galactic plane, and the coordinate system is right-handed, the eigenvalues of the tensor are $\lambda_x = -\partial^2\phi_G/\partial x^2 + \partial^2\phi_G/\partial z^2$, $\lambda_y = 0$, and $\lambda_z = -\partial^2\phi_G/\partial z^2$.

The potential ϕ has five extrema or saddle points, known as the Lagrange points; of these, by convention the points L_1 and L_2 lie at positions $L_{1,2} = (\mp r_{ti}, 0, 0)$, where

$$r_{ti} = \left(\frac{GM}{\lambda_x}\right)^{1/3}$$ (20)

is known as the tidal radius. At these points, the acceleration due to the gravity of the cluster is equal to that due to the combination of galactic gravity and the centrifugal force. The value of the potential at L_1 and L_2 is the Jacobi energy $E_J = -(3/2)GM_c/r_{ti}$, and the locus where $\phi(x, y, z) = E_J$ is called the Jacobi surface (Figure 15).

Stars with total energy $E < E_J$ are unable to cross the Jacobi surface, so any such stars located within it are bound to the cluster. Since the volume within this surface defines the volume of the star cluster, tidal truncation therefore defines a characteristic density $\rho_{ti} \approx 3M/4\pi r_{ti}^3 = 3\lambda_x/4\pi G$, which depends only on the tidal tensor and thus on the cluster's orbit and environment, and not on the cluster mass. Stars with energies $E \geq E_J$ are not bound to the cluster, but if they are located inside the Jacobi surface then they can only escape through a "window" around L_1 and L_2, with the size of the window depending on E/E_J. We show an example escape window for $E/E_J = 0.95$ (i.e., for a star with energy 5% greater than the Jacobi energy) in Figure 15. With the exception of those ejected during violent tidal shocks, most stars escaping a cluster have energies only slightly larger than E_J, so escape tends to be primarily near the Lagrange points.

4.2. Internal evolutionary processes

With these preliminaries understood, we can now ask what mechanisms can alter clusters' properties during the gas-free phase of their evolution.

4.2.1. Stellar evolution. At ages from $\approx 3 - 40$ Myr, stellar populations eject $\approx 20\%$ of their mass in SNe, which produce high-speed gas that almost certainly escapes the parent cluster; moreover, the remnant neutron stars or black holes left behind by SNe receive kicks of several hundred km s^{-1} (e.g., Faucher-Giguère & Kaspi 2006), so most escape the star cluster as well. After ≈ 40 Myr, SNe cease and mass loss

---

younger clusters.

[18]Because $\lambda_{x,y,z}$ are eigenvalues that arise from derivatives of the gravitational plus centrifugal tensor, Renaud, Gieles & Boily (2011) refer to them as the effective eigenvalues, to distinguish them from the eigenvalues of the tidal tensor by itself, computed without including the centrifugal potential.

becomes dominated by asymptotic giant branch (AGB) stars shedding their envelopes. This process is more gradual, so that the total mass loss reaches $\approx 40\%$ only after ≈ 1 Gyr, and does not reach 50% until ages of order a Hubble time. Moreover, while AGB wind velocities are high enough (tens of km s^{-1}) that little of the wind mass is likely to be retained, the white dwarfs left behind receive much smaller kicks, and thus are likely to remain in the cluster (e.g., Kruijssen 2009).

By itself a factor of two decrease in mass would be a small effect. However, the effects of mass loss can be greatly amplified by a tidal field. Adiabatic mass loss makes the cluster radius expand while the tidal radius shrinks; stars that find themselves outside $r_{\rm ti}$ as a result are lost. For clusters that begin their evolution filling their tidal radii, this effect roughly doubles the mass loss rate due to stellar evolution alone (Takahashi & Portegies Zwart 2000; Baumgardt & Makino 2003; Lamers, Baumgardt & Gieles 2010). Moreover, in the presence of a tidal field, for any specified cluster density profile there is a maximum ratio of half-mass radius to tidal radius, $r_h/r_{\rm ti}$, for which virial equilibrium is possible (Chernoff & Weinberg 1990; Fukushige & Heggie 1995). If mass loss drives $r_h/r_{\rm ti}$ above this maximum, the stars cannot find a new virial equilibrium, and the entire cluster disrupts. Both the additional mass loss and the propensity to catastrophic disruption are enhanced in mass-segregated clusters, where the stars losing mass carry a disproportionate share of the potential energy (Vesperini, McMillan & Portegies Zwart 2009; Gieles et al. 2010). The importance of stellar evolution depends on a star cluster's initial concentration and ratio of truncation radius to tidal radius $r_{\rm tr}/r_{\rm ti}$. Clusters with $r_{\rm tr}/r_{\rm ti} \ll 1$ or $W_0 \gtrsim 5$ ($c \gtrsim 1$) are dense enough to behave qualitatively like clusters that are not in a tidal field. Clusters with low concentration that already fill their tidal radii at the start of gas-free evolution are much more vulnerable.

Unfortunately there have been only limited observational studies of c and $r_{\rm tr}/r_{\rm ti}$ for young clusters. Ryon et al. (2015, 2017) survey massive ($> 5000\ M_\odot$) clusters in M83, NGC 628, and NGC 1313, and find that about half have $r_h/r_{\rm ti} \lesssim 0.15$ independent of cluster mass, small enough to render mass loss unimportant. The fraction with larger values of $r_h/r_{\rm ti}$ increases with age, as previously observed by Elson, Freeman & Lauer (1989) in the LMC, but too late for stellar mass loss to be effective. However, we caution that accurate measurements of r_h require accurate photometry in clusters' extended, low surface brightness halos, which is very difficult to obtain for galaxies at the distances of Ryon et al.'s sample. The result that the concentration does not correlate with mass is more robust, and appears to hold for OCs in the Milky Way (based on our examination of Kharchenko et al. 2013's sample) and the Magellanic Clouds as well (Fall & Chandar 2012, using the measurements of McLaughlin & van der Marel 2005), where measurements are easier. For Milky Way GCs, McLaughlin (2000) reports a weak correlation between concentration and mass, but no such correlation is apparent in the more recent fits of Baumgardt & Hilker (2018). Thus while there are hints that clusters might be born too concentrated for mass loss to be a significant effect, the observational case is far from settled. Moreover, we caution that all observations are potentially subject to both selection bias in favor of more compact clusters and survival bias – a substantial fraction of gas-free clusters at an age of ≈ 10 Myr might be only weakly bound and have $W_0 \approx 2 - 3$, for example, but these would be dramatically under-represented in OC catalogs because almost none would survive to reach ages more than a few tens of Myr.

4.2.2. Relaxation-driven mass loss. Stellar evolution has its greatest effects in the first $\sim 10 - 100$ Myr of a cluster's lifetime. On much longer time scales, the dominant internal process is mass loss driven by relaxation, sometimes also referred to as dissolution or evaporation. The underlying mechanism is simple: because stars with too much energy escape from clusters, the high-energy tail of a cluster's velocity distribution is underpopulated compared to a Maxwellian. As stars randomly exchange energy via gravitational interactions, the velocity distribution attempts to relax to a Maxwellian, repopulating the tail. Once stars scatter to high enough energy by this process, however, they escape through the windows around L_1 and

L_2, leaving the tail again underpopulated and continuing the cycle. Throughout the Galactic halo, we see direct evidence of this process through the disruption of GCs, e.g., Pal 5 (Odenkirchen et al. 1997, 2002), NGC 5466 (Grillmair & Johnson 2006), and Pal 14 (Sollima et al. 2011). Stars escaping through L_1 are at smaller galactocentric radii and thus move faster in the galaxy than the cluster which gives rise to a leading stream; stars escaping through L_2 are slower than the cluster and therefore trail. The escaping stars give rise to spectacular, cold stellar streams extending up to 60° across the sky (Odenkirchen et al. 2002; Grillmair & Johnson 2006). Limited evidence for the tidal disruption of OCs has been claimed but is rather less convincing (e.g., Berkeley 17; Bhattacharya et al. 2017).

Since this mass loss process is driven by repopulation of the Maxwellian tail as the velocity distribution relaxes, the loss rate scales approximately as $\dot{M}_{\rm rlx} \propto M/t_{\rm rlx}$, where $t_{\rm rlx} \approx 0.1(N/\ln N)(r_h/\sigma)$ is the relaxation time at the half-mass radius, M is the cluster mass, N is the number of stars in the cluster, and σ is the velocity dispersion. Hénon (1961, 1965a,b) carried out some of the earliest modeling of this process, finding that $\dot{M}_{\rm rlx} = f_{\rm rlx}M/t_{\rm rlx}$ with $f_{\rm rlx} \approx 0.05$, i.e., clusters lose about 5% of their mass per relaxation time. If we set $N = M/\overline{m}$, where $\overline{m}$ is the mean stellar mass, and $\sigma \approx \sqrt{GM/r_h}$, the mass loss rate scales as $\dot{M}_{\rm rlx} \propto \sqrt{\rho}$, where $\rho \propto M/r_h^3$ and we have dropped the weak mass dependence in $\ln N$. If clusters have density independent of mass (requiring r_h vs. M only a slightly steeper than the relation in Figure 9), or if they are tidally limited and thus have $\rho \approx \rho_{\rm ti}$ independent of M (§ 4.1), then $\dot{M}_{\rm rlx}$ is independent of M as well. This process is the justification for the "evolved Schechter" function form introduced in § 2.2 to fit the globular CMF, and explains how a Schechter-like ICMF could evolve into such a distribution (Fall & Zhang 2001; McLaughlin & Fall 2008).

While $\dot{M}_{\rm rlx} \propto M/t_{\rm rlx}$ is a reasonable first approximation, N-body simulations suggest a slightly more complex dependence. Fukushige & Heggie (2000), Baumgardt (2001) and Baumgardt & Makino (2003) show that, due to the finite time required for potential escapers to find the L_1 or L_2 windows and pass through them, the mass loss timescale depends on the crossing time $t_{\rm cr} = r_h/\sigma$ as well as the relaxation time, so $\dot{M}_{\rm rlx} \approx f_{\rm rlx}M/(t_{\rm rlx}^p t_{\rm cr}^{1-p})$, where $p \approx 0.7 - 0.8$ depending on W_0. The corresponding cluster dissolution timescale is

$$t_{\rm dis} = \frac{M}{\dot{M}_{\rm rlx}} \approx \begin{cases} 1.6\rho_2^{-1/2}M_3 \, {\rm Gyr}, & p = 1 \\ 0.6\rho_2^{-1/2}M_3^{0.7} \, {\rm Gyr}, & p = 0.7 \end{cases}, \tag{21}$$

where $\rho_2 = \rho/(100 \ M_\odot \ {\rm pc}^{-3})$, $M_3 = M/(10^3 \ M_\odot)$, and for the numerical evaluation we have used $f_{\rm rlx} = 0.05$, $\overline{m} = 0.5 \ M_\odot$, and $\ln N = 7.6$. Subsequent studies have shown that, in addition to W_0, p depends on the extent to which clusters initially fill their tidal radii (Gieles & Baumgardt 2008; Madrid et al. 2017), the ratios of the eigenvalues of the tidal tensor (Tanikawa & Fukushige 2010; Madrid et al. 2017), and the mass spectrum and degree of mass segregation in the cluster (Hong et al. 2013). Goudfrooij & Fall (2016) argue that p should also depend on cluster mass, and that the observed GC luminosity function implies $p \approx 1$ for globulars; however, this analysis does not include possible mass-to-light ratio effects due to correlations between GC metallicity and formation history.

A number of authors have proposed analytic models for cluster evolution that attempt to capture the findings of the N-body simulations, though none thus far have been able to capture the full range of numerical results over all parameter space. Lamers et al. (2005) and Lamers, Baumgardt & Gieles (2010) propose a model for cluster evolution incorporating the dependence of mass loss on the crossing time and the initial tidal radius, which gives $\dot{M}_{\rm rlx} \approx M^{1-\gamma}/t_0$, with $\gamma \approx 0.65$ and $t_0 \propto 1/\Omega$, where Ω is the angular velocity of the cluster's orbit through its host galaxy. Gieles, Heggie & Zhao (2011) propose a pair of "unified evolution equations" that capture the process of expansion and subsequent contraction that occurs when a cluster born with $r_{\rm ti} < r_{\rm tr}$ expands to its tidal radius and then shrinks as mass loss reduces that

radius:

$$\xi \equiv -\frac{\dot{N}}{N} t_{\mathrm{rlx}} = \frac{3}{5} \zeta \frac{t_{\mathrm{cr}}}{t_{\mathrm{cr},1}} \qquad\qquad \chi \equiv \frac{\dot{t}_{\mathrm{cr}}}{t_{\mathrm{cr}}} t_{\mathrm{rlx}} = \frac{3}{2} \zeta \left(1 - \frac{t_{\mathrm{cr}}}{t_{\mathrm{cr},1}} \right), \tag{22}$$

where ξ and χ are dimensionless rates of mass loss and cluster expansion and $t_{\mathrm{cr},1}$ is the crossing time of the cluster at the point where it has expanded to its tidal radius. The quantity ζ is the fraction of the total energy conducted per t_{rlx}, and has a complex dependence on the stellar mass spectrum and on the cluster core properties that must be calibrated by N-body simulations. The idea that clusters are born with $r_{\mathrm{tr}} < r_{\mathrm{ti}}$ and expand during their first ≈ 200 Myr of life is consistent with the tentative findings of Ryon et al. (2015, 2017), and with the observation that a higher proportion of the older GC population, though not all, fill their tidal radii (Innanen, Harris & Webbink 1983; Baumgardt et al. 2010).

4.3. External and environmental processes

Our discussion of internal processes implicitly assumed a constant tidal field. However, as clusters orbit the tidal field they experience may change due to cluster encounters with massive objects, or because the cluster's orbit is not a simple circle in the galactic plane. Such changes in the tidal potential cause the Jacobi surface to move, and sudden perturbations can also alter stellar energies. Both effects can move stars across the Jacobi surface, thereby unbinding part of a cluster. We now explore processes of this type.

4.3.1. Tidal perturbations. For star clusters with near-circular orbits in a galactic plane, the primary source of tidal perturbations is encounters with GMCs (Spitzer 1958). If an encounter takes place over a time shorter than the crossing time of a cluster, one can approximate it as impulsive (e.g., Binney & Tremaine 2008). Formally, this requires that the encounter velocity obey $v \gtrsim b\sqrt{GM/r_h^3} = 2.1 b_0 r_{h,0}^{-3/2} M_3^{1/2}$ km s^{-1}, where b is the impact parameter, $b_0 = b/1$ pc, M is the cluster mass, $M_3 = M/10^3 \; M_\odot$, and r_h is the cluster half-mass radius, $r_{h,0} = r_h/1$ pc. Since the typical encounter velocity in the ISM will be of order the ISM velocity dispersion (≈ 10 km s^{-1} in local galaxies), encounters between GMCs and all but the most massive clusters tend to be impulsive. Impulsive encounters can be further classified into catastrophic ones that completely disrupt the cluster, and diffusive ones that do not. Catastrophic disruption occurs only for perturbers that are denser than the cluster, while most star clusters are at least somewhat denser than Milky Way GMCs, $n \approx 100$ cm^{-3} (c.f. Figure 9), so the diffusive regime is more common.

The specific energy added per tidal shock in the diffusive regime, and the resulting mass loss rate, has been discussed by many authors (e.g., Gnedin & Ostriker 1999; Gieles, Athanassoula & Portegies Zwart 2007; Prieto & Gnedin 2008; Binney & Tremaine 2008). For the case of shocking by GMCs of finite size, Gieles et al. (2006) obtain the approximate relations

$$\frac{\Delta E}{E} \approx f(\tilde{b}) \left(\frac{M_{\mathrm{GMC}}}{M} \right)^2 \left(\frac{\sigma}{v_{\mathrm{max}}} \right)^2 \qquad\qquad \frac{\Delta M}{M} \approx f_M \frac{\Delta E}{E}, \tag{23}$$

where E and M are the cluster total energy and mass, ΔE and ΔM are the changes in these quantities due to the encounter (with a sign convention whereby $\Delta M > 0$ corresponds to mass loss), $f_M \approx 0.25$ is the ratio of fractional mass loss to fractional energy gain (less than unity since stars escape with finite energy), $\sigma = \sqrt{GM/r_h}$ is the stellar velocity dispersion in a cluster, v_{max} is the maximum velocity during the encounter, and $f(\tilde{b})$ is a numerical factor that depends on the dimensionless impact parameter $\tilde{b} \equiv b/r_h$ as $f(\tilde{b}) \to$ const for $\tilde{b} \ll 1$ and $f(\tilde{b}) \to \tilde{b}^{-4}$ for $\tilde{b} \gg 1$. The time to disrupt a cluster, $t_{\mathrm{dis}} = M/\dot{M}$, can be found by integrating over encounters with a population of GMCs. Second-order energy injection (Kundic & Ostriker 1995) reduces the Gieles et al. (2006) disruption time by a factor $C_{\mathrm{sh}} \simeq 0.3$ (Kruijssen et al.

2011) so that the disruption time for a cluster of mass M becomes

$$t_{\mathrm{dis}} \approx \frac{C_{\mathrm{sh}}}{\sqrt{\pi}} f_{\mathrm{dis}} \frac{\sigma_{\mathrm{ISM}}}{G\Sigma_{\mathrm{GMC}}} \left(\frac{\rho}{\rho_{\mathrm{GMC}}} \right) \approx 1.6 \sigma_1 \Sigma_2^{-1} \left(\frac{\rho/\rho_{\mathrm{GMC}}}{1000} \right) \text{ Gyr} \tag{24}$$

where σ_{ISM} is the velocity dispersion of the ISM (assumed equal to the velocity dispersion of the clusters), Σ_{GMC} is the surface density of an individual GMC, ρ_{GMC} is the mass density of GMCs in the ISM (not the GMC internal density), $\rho = 3M/8\pi r_h^3$ is the cluster density, and f_{dis} is a factor of order unity that depends on gravitational focusing and cluster structure; for Gieles et al. (2006)'s favored parameters, $f_{\mathrm{dis}} \approx 0.4$. We have scaled to Milky Way values (c.f. § 3.1.1) for the other parameters: $\sigma_1 = \sigma_{\mathrm{ISM}}/10$ km s$^{-1} \approx 1$, $\Sigma_2 = \Sigma_{\mathrm{GMC}}/100$ M$_\odot$ pc$^{-2} \approx 1$, and $\rho/\rho_{\mathrm{GMC}} \approx 1000$. The 1.6 Gyr disruption timescale is substantially longer than earlier estimates (e.g., Binney & Tremaine 2008), largely because Gieles et al. (2006) find $f_M \approx 0.25$, while some earlier analytic models adopted $f_M \approx 1$.

One factor that could increase the disruption time is that tidal shocks tend to be self-limiting: a first tidal shock strips off the most weakly bound stars, but the timescale for the cluster to repopulate the high energy orbits these stars occupied is dictated by two-body relaxation (§ 4.2.2), so that subsequent tidal shocks separated in time by less than t_{rlx} remove much less mass (Kruijssen 2015; Gieles & Renaud 2016). On the other hand, the numerical evaluation in Equation 24 could underestimate the importance of tidal heating because it is based on mean conditions in the ISM. However, star clusters are formed in regions of higher than average gas density and structure, i.e., larger ρ_{GMC} and Σ_{GMC}. Star cluster locations remain correlated out to ages ≈ 100 Myr (Grasha et al. 2017a), and during this time they likely experience enhanced tidal perturbations, a phenomenon that Kruijssen et al. (2011) dubbed the "cruel cradle effect". Several authors have suggested this effect may explain why cluster number counts decline from $10-100$ Myr, when no other mechanism except stellar mass loss (§ 4.2.1) is expected to operate (Elmegreen & Hunter 2010; Kruijssen 2012; Miholics, Kruijssen & Sills 2017). It is likely to be particularly important in gas-rich environments such as high-redshift galaxies and local mergers, and may explain why globular clusters have a lognormal rather than a power-law mass function: since $t_{\mathrm{dis}} \propto \rho$, if density increases with mass then low-mass clusters will be preferentially destroyed by tides, while massive clusters will survive (Elmegreen 2010; Kruijssen 2015). However, the importance of this effect depends on the assumed mass-radius relation. Most published models assume mass-independent cluster radii, so $\rho \propto M$, which tends to strengthen the preferential destruction of low-mass clusters compared to the mass-radius relation shown in § 2.5.

This situation is somewhat different for clusters in disk-crossing or bulge-crossing orbits, which experience changing tides as they pass through a disk or bulge. During each passage the galaxy exerts a pinch compression on the cluster that acts normal to the disk. The formalism for modelling the effects of a disk passage is largely the same as that for GMC encounters, simply with a different frequency and shock strength, and with some subtle modifications for the fact that GMC tidal shocks are 2D, while disk tidal shocks are 1D (Gnedin & Ostriker 1997). Starting with Ostriker, Spitzer & Chevalier (1972), there is an extensive literature on transiting clusters, with an in-depth review provided by Binney & Tremaine (2008). The timescale for complete destruction of globulars through most of the Milky Way disk is much longer than a Hubble time, but erosion of the outer stars requires only ≈ 10 Gyr. Thus, the stellar escape feeding the cold streams observed in Pal 5 (Odenkirchen et al. 1997, 2002), for example, may have been exacerbated by repeated crossings. Globulars that transit the bulge of a galaxy are subjected to much stronger tides, and may lose substantial fractions of their mass in well under a Hubble time (Madrid et al. 2017).

For open clusters, which are less massive and dense, disk crossing is more perilous. Martinez-Medina et al. (2017, see also Webb et al. 2014) find that the destruction rate is nearly independent of vertical velocity ($\propto v_z^{-0.2}$) in conflict with the v_z^{-2} dependence found by Ostriker, Spitzer & Chevalier (1972); Martinez-Medina et al. find the number of disk crossings is more important than the velocity. Clusters far from the plane transit the disk only rarely, whereas clusters close to the plane experience a weaker gradient

when they do transit. Consequently, clusters a few hundred parsecs from the plane are the most vulnerable to destruction by tidal shocking. This may explain the clear dichotomy of OCs close to the plane and GCs far from the plane, with very little crossover in phase space. Interestingly, the vast majority of reported OCs more than a few hundred parsecs off the plane are not confirmed in the latest *Gaia* DR2 analysis (Cantat-Gaudin et al. 2018).

4.3.2. Migration and environment. Several mechanisms alter the orbits of star clusters as they evolve, potentially shaping their demographics. Here we discuss processes that move clusters within a single galaxy, deferring discussion of galaxy-galaxy interaction to § 4.4.2. One potentially important mechanism is dynamical friction, which tends to move clusters toward galactic centers by causing them to lose angular momentum. For a cluster of mass M moving with velocity $\mathbf{v}_c$ through an inertial stellar field described by a distribution function $f(v)$ for stars with mass m and velocity v, Chandrasekhar (1943) showed

$$\frac{d\mathbf{v}_c}{dt} = -16\pi^2 G \ln \Lambda \, \frac{m(m+M)}{v_c^3} \int_0^{v_c} v^2 f(v) \, dv \, \mathbf{v}_c \tag{25}$$

for which $\ln \Lambda$ is the Coulomb logarithm; this expression can be simplified considerably if $f(v)$ is a Maxwellian. Massive clusters decelerate more rapidly than lower mass ones ($d\mathbf{v}_c/dt \propto M$ for $M \gg m$), because they create larger wakes. This is one of the few mechanisms that encourages the destruction of the most massive clusters: Bekki (2010) finds that dynamical friction causes clusters with masses $\gtrsim 2 \times 10^5$ $M_\odot$ in the disks of sub-L_* galaxies to sink to the galactic center in less than a Hubble time, where stronger tides are likely to destroy them (Gerhard 2001). The effect is smaller in larger galaxies like the Milky Way.

Radial migration driven by transient spiral density waves, first discussed by Sellwood & Binney (2002) and demonstrated numerically by Roškar et al. (2008), can move stars inward or outward. The same mechanism works for clusters, and is non-destructive to them since clusters have internal crossing times much shorter than the time required for them to transit a spiral density wave (Gieles, Athanassoula & Portegies Zwart 2007). Migration occurs because a spiral perturbation with pattern speed Ω_P can modify an object's energy E and angular momentum J while conserving its Jacobi integral I_J; in the (E, J) plane, objects move along lines of constant $I_J = E - \Omega_P J$. A single spiral wave near co-rotation can perturb the angular momentum by 20% without significant heating, moving objects from one circular orbit to another, inwards or outwards. Multiple transient spirals lead to a random walk in the (E, J) plane at constant I_J (Dehnen 2000), so that substantial variations in the angular momentum of an object can occur over Gyr timescales. By contrast, long-term spiral arms produce no net effect, because perturbations as an object enters an arm are cancelled by those as it exits (Lynden-Bell & Kalnajs 1972). Objects that oscillate far from the plane or are on highly radial orbits are much less influenced by the migration mechanism. Minchev & Famaey (2010) and Minchev et al. (2011) propose a related mechanism that does not rely on transience: quasi-chaotic interference between resonances from multiple rotating patterns, e.g., the bar and the spiral arms, with different pattern speeds (c.f. Brunetti, Chiappini & Pfenniger 2011). Jílková et al. (2012) and Quillen et al. (2018b) investigate this resonance overlap and find that outward migration is possible but relatively inefficient. External influences can also drive radial migration, in particular, radial in-plane orbiting galaxies that come close enough to strongly perturb the disk (Quillen et al. 2009).

Minchev & Famaey (2010) provide a comprehensive discussion of the evidence for strong migration of individual stars over billions of years, in particular, the ∼1 dex scatter in [Fe/H] at all radii. The high local fraction of metal-rich stars may also be evidence of migration (Kordopatis et al. 2015), since the local ISM has somewhat sub-solar metallicity (Nieva & Przybilla 2012). The case for migrating OCs is less clear. There are at least half a dozen supersolar metallicity OCs within ∼ 1 kpc of the Solar Circle (Quillen et al. 2018b). Moreover, Friel (1995) finds that there is a large scatter in the metallicity and age of OCs at all radii.

However, given the lower statistics, it is unclear whether this provides compelling evidence for cluster migration within the disk, particularly since there is an alternative possibility: OCs that seem out of place with respect to the average properties of their environment may have been accreted from infalling dwarfs. Carraro & Bensby (2009) argue that the OCs Berkeley 29 and Saurer 1 appear to be associated with the disrupting Sgr dwarf galaxy, and the case for some Milky Way GCs having arrived with Sgr is compelling: *Gaia* has shown that four GCs (Arp 2, Terzan 7, Terzan 8, M54) confined to the Galaxy's central core are co-moving (q.v. Vasiliev 2018). An OC that is arguably an even stronger candidate for having arrived by infall rather than migration is NGC 6791, an old (~ 8.3 Gyr), massive ($\sim 4000\ M_\odot$), extremely metal-rich ([Fe/H] $\approx +0.3$) OC along the Solar Circle (Boesgaard, Lum & Deliyannis 2015), 4 kpc from the Sun and 0.8 kpc above the plane (Carraro et al. 2006; Origlia et al. 2006). Its large vertical distance argues against radial migration from the bulge region. While a likely host for infall has not been identified to date, *Gaia* DR2 data has begun to provide evidence that massive accreted systems like Sgr are dispersed throughout the Galaxy (e.g., Gaia-Enceladus; Helmi et al. 2018b). We anticipate that the more comprehensive *Gaia* data release (DR3) in 2020 will significantly clarify whether infall can account for all the anomalous OCs, or whether some must be the result of migration.

4.4. Models of the full cluster population

Having reviewed the individual processes affecting gas-free clusters, we now discuss attempts to combine them with cluster formation into models for the complete cluster population. We first consider clusters confined to galactic disks, and second examine clusters outside of disks.

4.4.1. Disk clusters. Comparison of the observed distribution of cluster masses and ages with theoretical models sheds light on the rate at which clusters are born and on the processes that destroy them. If the mass of a cluster is determined by its initial mass, M_i, and its age, T, then the number of surviving clusters in a given age range is the same as that when they were born, so that (Fall & Zhang 2001; Lamers et al. 2005; Fall, Chandar & Whitmore 2009)

$$\frac{\partial^2 N}{\partial M\,\partial T} = C(T) M_i^{\alpha_{Mi}} \left(\frac{\partial M_i}{\partial M}\right)_T, \tag{26}$$

where $C(T)$ is proportional to the birthrate of clusters with a current age T and where we have assumed that the ICMF is a power law in the range of masses considered; as discussed in § 2.2, $\alpha_{Mi} \approx -2$. Insofar as the effects of stellar evolution occur prior to relaxation (a good approximation) or tidal effects (a good approximation after the cluster has left its natal complex), one can define $M_0 = \mu_{\rm ev}(T) M_i$ as the mass at the onset of relaxation and tidal effects (Lamers et al. 2005), where $\mu_{\rm ev}(T)$ drops from unity to $\sim 0.25 - 0.5$ during the first few 10^7 years of the cluster's life.

Two classes of models for cluster evolution have been considered, mass-independent destruction (MID; Fall, Chandar & Whitmore 2005; Whitmore, Chandar & Fall 2007; Fall, Chandar & Whitmore 2009) and mass-dependent destruction (MDD; Boutloukos & Lamers 2003; Lamers et al. 2005; Lamers, Baumgardt & Gieles 2010). For the MID case, the cluster distribution function breaks up into the product of a mass-dependent term and an age-dependent term, which are taken as power laws, so that

$$\frac{d^2 N}{dM\,dT} \propto M^{\alpha_{Mi}} T^{\alpha_T}. \tag{27}$$

Since the effects of stellar evolution are approximately independent of mass as long as initial concentration c does not depend on mass, its effects are naturally included in this formulation. This model is phenomenological, since the rate of cluster destruction is not calculated.

For the MDD case, the destruction time is generally taken to be a power-law function of mass (Boutloukos & Lamers 2003; Lamers et al. 2005), $dM/dt = -M/(t_0 M^\gamma)$, where M is measured in solar masses. If t_0 is constant, then (Lamers et al. 2005)

$$\frac{\partial^2 N}{\partial M \, \partial T} = \frac{C(T)\mu_{\rm ev}(T)M^{\alpha_{Mi}}}{[1 + \gamma T/(t_0 M^\gamma)]^{(\gamma+1)/\gamma}}, \tag{28}$$

Values of γ estimated from observation (e.g., Boutloukos & Lamers 2003; Gieles 2009) and N-body simulations (Lamers et al. 2005; Lamers, Baumgardt & Gieles 2010) are usually in the range $0.6 - 0.7$ (see § 4.2.2). At late times or low masses, the second term in the denominator dominates, so that the distribution function becomes proportional to $M^{\alpha_{Mi}+\gamma}T^{-(\gamma+1)/\gamma}$. However, because clusters fade with time, it is sometimes the case that in order to be complete, cluster samples must be constructed so that the minimum mass is large enough that the second term is not dominant. It is not clear that any extragalactic sample of clusters has sufficient dynamic range to clearly show the break in the slope of the mass function.

There are two complications with the MDD model. First, t_0 may not be constant: clusters move from a dense environment where they are born to the normal ISM over a period that can extend to 100 Myr (and much longer in galaxy mergers), and over this time t_0 is likely to increase (Elmegreen & Hunter 2010; Kruijssen et al. 2011). Second, the simulations used to calibrate this model (e.g., Baumgardt & Makino 2003; Lamers, Baumgardt & Gieles 2010) have included only relaxation, not tidal losses due to encounters with GMCs, yet in the Galaxy the latter process is likely dominant. The characteristic destruction time for clusters due to GMCs is a few hundred Myr (§ 4.3.1), while for a $10^4 \, M_\odot$ cluster the disruption time due to relaxation (confirmed by simulations – Lamers et al. 2005) is multiple Gyr (Equation 21). The time scale for destruction by encounters with GMCs depends only on the density of the cluster, so agreement between the observed and simulated values of γ occurs only if $\rho \propto M^{0.6-0.7}$ – possible given the uncertainties in Figure 9, but by no means certain.

Observational evidence has been marshalled on both sides of this debate, which is closely tied to the issue of inclusive versus exclusive catalog construction discussed in § 2.1.2 and § 2.3. Equation 27 predicts a cluster age distribution with constant α_T, and thus is "universal", while Equation 28 predicts a value of α_T that smoothly varies from 0, for $\gamma T/(t_0 M^\gamma) \ll 1$, to $-(\gamma+1)/\gamma$, for the opposite limit. Thus authors who construct exclusive catalogs, which tend to produce α_T close to zero at early ages and then lower values at older ages, generally favor MDD models. Authors who construct inclusive catalogs, which generally have α_T close to -1 at all ages, tend to favor MID models.

In the latter category, Whitmore, Chandar & Fall (2007) and Fall, Chandar & Whitmore (2009) show that the data on clusters in the Antennae are well described by a power law of the form in Equation 27 and suggest that this is a "quasi-universal law" of cluster destruction. Fall & Chandar (2012) extended this analysis to clusters in several other galaxies, including the SMC, the LMC, and M83; in all cases, dN/dM and dN/dT exhibit power law behavior with indexes $\alpha_M \approx -2$ and $\alpha_T \approx -0.8$. However, for Milky Way OCs, $\alpha_T \simeq -0.6$ to -0.5 over the age range $10 - 300$ Myr (Figure 6), inconsistent with universality. Fall & Chandar 2012 also search for deviations from power-law behavior in the LMC as predicted by the MDD model, by finding clusters with both $\gamma T > t_0 M^\gamma$ and $\gamma T < t_0 M^\gamma$. They find that simple MDD models fail for the LMC. However, they did not include stellar mass loss, which is often included in MDD models (e.g., Lamers, Baumgardt & Gieles 2010), allow for a time-dependent destruction time (e.g., Elmegreen & Hunter 2010), or consider the possibility that some of the youngest clusters may be unbound and therefore dissipate rapidly.

Supporting the MDD picture, Lamers et al. (2005) show that the cluster population within 600 pc of the Sun (Kharchenko et al. 2005) is consistent with dN/dT calculated from Equation 28 with $\gamma = 0.62$ and a 1.0 Gyr destruction time for $10^4 \, M_\odot$ clusters; the data cannot be fit by a power law. Bastian et al. (2012)

generalize the cluster evolution model to include both a mass-dependent and mass-independent destruction term, and find that MID and MDD models give equally good fits to two fields in M83. Adamo et al. (2017) find evidence for mass-dependent destruction of clusters with $M < 10^4 \, M_\odot$ in NGC 628, and Messa et al. (2018b) report a similar result for M51. In both cases, however, the evidence comes from the lowest mass bin, so deeper observations or statistical methods for making better use of partially-complete data (e.g., Krumholz et al. 2018) would be beneficial.

Theoretically, the key difference between the two models lies in the mass dependence of the cluster disruption time, $t_{\rm dis}$. As discussed in § 4.3.1, the dominant destruction process for disk clusters is likely tidal shocking by GMCs, yielding disruption time proportional to the cluster density, $t_{\rm dis} \propto \rho$. If ρ is approximately independent of mass, then so is $t_{\rm dis}$ and one gets the MID model. Mass-independent densities are in fact expected for tidally limited clusters (Fall & Chandar 2012), but the fact that observed young clusters have EFF rather than King profiles (§ 4.1) and (at least potentially) underfill their tidal radii (Ryon et al. 2015, 2017) suggests they have not yet reached equilibrium. On the other hand, advocates of the MDD model (e.g., Lamers et al. 2005) assume $\rho \propto M^{0.6-0.7}$, so that the mass dependence of tidal disruption is similar to that for 2-body relaxation. Their models can therefore implicitly include tidal disruption by GMCs, although not with the correct rate. Unfortunately, cluster densities are difficult to determine (§ 2.5), so to date it has not been possible to use measurements of ρ vs. M to distinguish the two models.

Miholics, Kruijssen & Sills (2017) carry out simulations of isolated galaxies using a subgrid model for the formation and evolution of star clusters (Kruijssen et al. 2011, 2012b); the model includes both relaxation and encounters with GMCs. They demonstrate that, in the MDD model, the median age of clusters in a sample is proportional to t_0, and the median age depends on the surface density and velocity dispersion of the gas. Comparing with observations of M31 (outer galaxy), M51 (three mass ranges), and M83 (inner and outer regions), they find the best fit for the mass dependence of the median age is $t_{\rm med} \propto M^{0.54}$. In all their models, dN/dT decreases with age even though the star formation rates are roughly constant, and they are able to fit the observed median ages for the six cluster samples to within 0.2 dex, ranging from $10^{7.3}$ yr for clusters in M51 with masses greater than $10^3 \, M_\odot$ to $10^{8.3}$ yr for clusters in the outer parts of M31 and M83. This more complete model for cluster evolution is more complex than the simple MID and MDD models discussed above, and it demonstrates that cluster destruction is environmentally dependent.

4.4.2. Non-disk clusters. All star clusters, including the progenitors of GCs, likely began their lives in the disk of some galaxy (though see Mandelker et al. 2018 for a model where this is not true). Thus clusters found outside galaxy disks necessarily have complex histories. Studies of these histories therefore focus on determining possible origin sites for the clusters we see today, on the processes by which those clusters reached their present locations. A great deal of this work has been devoted to modeling the cosmological assembly history of the Milky Way's GC system, and to explaining observed features such as its color bimodality and color-magnitude relationships (the "blue tilt"; Brodie & Strader 2006). These topics are beyond the scope of our review, which we limit to processes affecting star clusters in general, rather than the specific histories of GCs.

Since the transfer of clusters out of disks involves strong tides, simulations of the origin of non-disk clusters must model tidal effects. Since galactic and cosmological simulations cannot resolve cluster interiors (§ 3.5.3), they cannot do so directly, and instead the usual approach is to treat each star cluster as a point mass that is either formed self-consistently in the simulation or inserted following some prescribed initial distribution. As the particle moves along its orbit, one records the tidal potential to which it is subject, and uses this to calculate cluster evolution from an analytic model (e.g., Prieto & Gnedin 2008; Kruijssen et al. 2011; Brockamp et al. 2014; Rossi & Hurley 2015b; Carlberg 2018; Pfeffer et al. 2018) or via a direct N-body simulation (e.g., Renaud & Gieles 2013; Rieder et al. 2013; Rossi & Hurley 2015a; Mamikonyan

et al. 2017). The latter approach is more accurate, but due to its expense one can only simulate a few examples, and for less than a full Hubble time.

One obvious potential site for the formation of non-disk clusters is in major galaxy mergers such as the Antennae. Renaud & Gieles (2013) and Kruijssen et al. (2012b) both simulate Antennae-like mergers, but reach opposite conclusions: Renaud & Gieles find that the merger has little effect on the cluster population, while Kruijssen et al. find that it transforms the CMF from Schechter-like to lognormal like. The Kruijssen et al. results are likely more reliable. They include gas and form clusters self-consistently during the merger, while Renaud & Gieles carry out pure N-body simulations and study only pre-existing clusters; both differences produce weaker tides in Renaud & Gieles's simulations. Renaud & Gieles also initialize their clusters with much smaller radii, making them more resistant to tidal disruption.

Dwarf galaxies represent another potential formation site for clusters that end up in the halos of larger galaxies such as the Milky Way: clusters form in the dwarf (and are therefore metal-poor, as observed), and then become part of the halo when the larger galaxy accretes and tidally strips the dwarf. As discussed in § 4.3.2, there are multiple Milky Way clusters that may have originated in this way. Simulations suggest that the accretion process is gentle, so accreted clusters do not experience major tidal losses, and quickly adjust to their new environments and become indistinguishable, except in their abundances, from the pre-existing cluster population (Miholics, Webb & Sills 2014, 2016; Bianchini et al. 2015).

Exploring the full GC population requires cosmological simulations. Unfortunately, it is not presently possible to use the same simulations to study both the formation of GCs and their subsequent evolution (see Forbes et al. 2018 for more discussion). Simulations that can resolve cluster formation (e.g., Kimm et al. 2016; Ricotti, Parry & Gnedin 2016; Li et al. 2017; Li, Gnedin & Gnedin 2018; Kim et al. 2018) are too expensive to run to redshift zero, so studies that reach the present use potential GC progenitors inserted by hand. Different authors make different assumptions about how this should be done. Prieto & Gnedin (2008) and Carlberg (2018) use dark matter-only simulations; the former initialize cluster particles based on an analytic model of the baryonic mass distribution, while the latter adopt a randomly-oriented disk. They also assume different redshift distributions. Renaud, Agertz & Gieles (2017) and Pfeffer et al. (2018) use dark matter plus baryon cosmological simulations, with Renaud, Agertz & Gieles assuming that clusters follow the distribution of all newly-formed stars and Pfeffer et al. using the analytic models of Kruijssen (2012) and Reina-Campos & Kruijssen (2017) to vary Γ and the ICMF depending on local ISM conditions.

The conclusions are as diverse as the sets of initial conditions: Prieto & Gnedin (2008) and Pfeffer et al. (2018) find that tides and two-body relaxation are able to transform an initially-power-law distribution of cluster masses into a lognormal, consistent with the observed globular CMF. Renaud, Agertz & Gieles (2017) reach the opposite conclusion, finding that their test clusters experience tides too weak to produce such a radical change. The difference is likely because their recipe for cluster formation – clusters simply follow star formation, in contrast to Prieto & Gnedin's prescribed distribution or Pfeffer et al.'s variable Γ and ICMF – places clusters on orbits where they experience weaker tides. Similarly, Carlberg (2018) finds that the effectiveness of tides is sensitive to the assumed distribution of redshifts for cluster formation. As for the ICMF (§ 3.5.3), at present simulation outcomes appear to depend primarily on subgrid recipes, and thus have little predictive power. Fortunately, we are likely to gain far better observational constraints once *JWST* launches, since it should narrow down the range of plausible models considerably.

5. AFTERLIFE

The clusters we observe today are the surviving remnant of a much larger population that have now dissolved. However, within the Milky Way it is possible to reconstruct some of these clusters via a variety

of techniques. Doing so provides important clues about the cluster formation and dispersal process, the Galactic potential and, for clusters in the halo, the nature of dark matter. We therefore focus this final section on cluster reconstruction using kinematic (§ 5.1), action-angle (§ 5.2), and chemical tagging methods (§ 5.3). This field is in its infancy, so the discussion is necessarily more speculative than that up to this point.

5.1. Kinematics

The traditional method for cluster reconstruction is to find "moving groups": clusters of stars in velocity space.[19] Clustering in velocity space is useful because unbound stars disperse in position on a timescale of a crossing time (a few Myr), but only spread out in velocity over timescales of a substantial fraction of their orbital period in the Galaxy (tens of Myr). This technique has a long and checkered history (Griffin 1998) with some early identified groups being questioned in later analysis (q.v. Taylor 2000), reminiscent of the "faux clusters" identified in *Gaia* DR2 (Cantat-Gaudin et al. 2018). Nonetheless, it remains a powerful method for finding disrupted clusters.

Within the disk, Dehnen (1998) first used the *Hipparcos* survey to identify kinematic substructure in the (U, V) plane for a very local sample ($D \lesssim 100$ pc); he identified known star clusters, but also discovered the previously-unknown Hercules stream. Bensby et al. (2007) showed that this structure is inhomogeneous in its abundances, and therefore more likely to be stars trapped in a resonance (e.g., Quillen et al. 2018a) than a disrupting cluster. Subsequent wavelet analysis of the *Hipparcos* survey combined with the RAVE survey identified many more local clumps in the (U, V) plane, although they have proven difficult to interpret (Antoja et al. 2012). While resonant streams are interesting, they are "false positives" from the standpoint of cluster reconstruction. One major challenge for kinematic searches is that the expected rate of such false positives is unknown. The 2D simulations of De Simone, Wu & Tremaine (2004) suggest it is high, but we suspect that it will be lower in a 3D simulation due to the extra degree of freedom. For objects that are real disrupted clusters, one can use dynamical traceback methods to determine the age and other properties of the original cluster (de Zeeuw et al. 1999; Riedel et al. 2017).

The halo has a great deal of substructure, but most known structures are associated with the disrupting Sgr dwarf (Helmi 2008). Exceptions include cold halo streams like those extending from the GC Pal 5 (§ 4.2.2). There are similar linear structures such as GD-1 (Grillmair & Johnson 2006) and the "Jet" stream (Jethwa et al. 2018) not associated with a surviving progenitor; these are likely clusters that have fully disrupted (e.g., Koposov, Rix & Hogg 2010). Some halo substructures of comparable mass have only been identified kinematically (Kepley et al. 2007) or, for an assumed Galactic potential, by converting from kinematic coordinates to energy-angular momentum (Helmi et al. 2017) or angular momentum-eccentricity space (Helmi et al. 2006). Since *Gaia* DR2, new studies (q.v. Antoja et al. 2018) reveal many more substructures in the (U, V) plane, but follow-up work to identify them has only just begun.

5.2. Action-angle space

Kinematic reconstruction fails at ages $\gtrsim 100$ Myr because differential acceleration along stars' orbits makes them disperse in velocity space. However, while velocities change as stars orbit, there are conserved quantities that do not, at least to the extent that we can approximate the Galactic potential as smooth and

---

[19]This space is usually described by the three velocity components (U, V, W), where U is toward the Galactic Center, V is the velocity in the Galactic plane along the direction of the Sun's orbit, and W is the velocity out of the plane, oriented so the coordinate system is right-handed. Velocities are measured relative to the Local Standard of Rest (Bland-Hawthorn & Gerhard 2016).

time-independent. In principle these quantities act as labels that are fixed when stars form, allowing reconstruction of clusters long after they have dispersed in velocity. This insight motivates the idea of cluster reconstruction in action-angle space $(\mathbf{J}, \boldsymbol{\theta})$, a set of canonical conjugate coordinates defined by (Binney & Tremaine 2008)

$$2\pi J_i = \oint \dot{x}_i \, dx_i = \frac{1}{\Omega_i} \int_0^{2\pi} (\dot{x})^2 \, d\theta_i \tag{29}$$

where x_i are Cartesian coordinates and Ω_i are the orbital frequencies. For each star, the equations of motion with respect to the Hamiltonian $\mathbf{H}$ are

$$\dot{\boldsymbol{\theta}} = \frac{\partial \mathbf{H}}{\partial \mathbf{J}} = \Omega(\mathbf{J}) \qquad\qquad \dot{\mathbf{J}} = -\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}} = 0 \tag{30}$$

An orbit can be traced with the solution $\theta_i(t) = \Omega_i(t) + \Omega_{i,0}$. For the inertial Galactic system, (R, ϕ, z) coordinates[20] are most often used (even for halo studies) such that $\boldsymbol{\theta} = (\theta_R, \theta_\phi, \theta_z)$ and $\mathbf{J} = (J_R, J_\phi, J_z)$ where $J_\phi \equiv L_z$, i.e., J_ϕ is the angular momentum with respect to the spin axis. A stellar orbit is described by the three actions $\mathbf{J}$. Orbits with $J_z = 0$ lie in the Galactic plane, orbits with $J_R = 0$ are circular, and we choose units so that a circular orbit at the Solar Circle has $J_\phi = 1$; circular orbits with $J_\phi < 1$ lie inside the Solar Circle. The orbits of all stars in the Galaxy are described by the distribution function $f(\mathbf{J})$.

To find clusters in action-angle space one must convert the observed positions and velocities of stars into actions. This requires knowledge of the Galactic potential Φ, which enters the Hamiltonian $\mathbf{H}$. We will not fully review efforts to measure the Galactic potential and the stellar density distribution, which must be determined simultaneously. Arguably, the best efforts to date are from the RAVE survey, which provides an explicit fit for the vertical stellar density distribution (Piffl et al. 2014) as opposed to the parameterized potentials used in earlier work (e.g. McMillan 2011). Moreover, Φ is at least weakly time-dependent. In addition to the outer warp, there is clear evidence near and far of large-scale corrugation waves propagating throughout the disk (e.g. Xu et al. 2015; Antoja et al. 2018). The spiral arms and bar provide additional challenges.

Even with the uncertainties, the action-angle approach reveals rich structure throughout the halo and Galactic disk. Sellwood (2010) mapped *Hipparcos* moving groups to action space and found that this improved the delineation of the Hyades stream, but that other structures were barely evident. With *Gaia* DR2, the situation has improved and over a much larger volume than *Hipparcos*. Trick, Coronado & Rix (2018) show that the highest density peaks in the (U, V) plane map to the highest density peaks in the action plane, specifically $(\sqrt{J_R}, J_\phi = L_z)$, but the mapping is not simple. They trace structure out to 1.5 kpc from the Sun and find much of it has low vertical action, J_z, suggestive of disk resonances that operate

---

[20]In this coordinate system, R is the distance from the Galactic Center, z is distance out of the Galactic plane, and the line from the Sun to Sgr A* defines $\phi = 0$.

most efficiently in the plane (Katz et al. 2018). The Helmi streams were discovered with a dozen or fewer stars using action analysis (Helmi 2008), structures that were invisible against the background in kinematic space. Arguably the most impressive action-based discovery to date is Gaia-Enceladus, the remnants of a massive dwarf spread throughout the inner Galaxy that was accreted long before Sgr (Helmi et al. 2018a). To date, no OCs or GCs (or their streams) have been discovered using actions, but these are early days. There are many clumps in this space that have yet to be followed up. Stars bunched in phase space are typically spread across the sky requiring extended observing programs over $1 - 2$ years.

5.3. Chemical tagging

Dynamical "invariants" like actions are useful labels for reconstruction, but they are not strictly conserved over a star's lifetime, e.g., when transiting the spiral arm or bar. These time-dependent features of the Galactic potential induce perturbations in the actions of the stars they affect; indeed, non-conservation of J_ϕ due to arms and bars is responsible for stellar migration ($\S$ 4.3.2), and the same structures also likely cause non-conservation of J_R (Solway, Sellwood & Schönrich 2012). This ultimately limits how far back in time action-angle reconstructions for clusters in the disk can go. Longer-term reconstructions therefore require another invariant, for which elemental abundances are a natural choice. Stars conserve their abundances over almost their entire lives, and stars in the same cluster have nearly-identical abundances ($\S$ 2.6). This led Freeman & Bland-Hawthorn (2002) to suggest that, with abundance data of sufficient quality and enough elements, clusters that have dispersed even in action space could be detected via clustering in $\mathcal{C}$-space, defined so each star is a point in a high dimensional space, i.e., $\mathcal{C}([\text{Fe/H}], [\alpha/\text{Fe}], [\text{X}_1/\text{Fe}], [\text{X}_2/\text{Fe}], \ldots)$.

Weak chemical tagging, defined as identifying distinct components separated by a surface in $\mathcal{C}$-space, has been used to good effect. For example, stars in the Solar neighborhood show a clear bimodality in $\mathcal{C}([\text{Fe/H}], [\alpha/\text{Fe}])$ (e.g., Bensby et al. 2005) and $\mathcal{C}([\alpha/\text{Fe}], [(\text{C+N})/\text{Fe}], [\text{Al/Fe}], [\text{Mg/Mn}])$ (Hawkins et al. 2015). The Sgr dwarf and stream separate cleanly from all components of the Milky Way in the high-dimensional space $\mathcal{C}([(\text{C+N})/\text{Fe}], [\text{O/Fe}], [\text{Mg/Fe}], [\text{Al/Fe}], [\text{Mn/Fe}], [\text{Ni/Fe}])$ (Hasselquist et al. 2017). Reconstructing star clusters, however, requires strong chemical tagging, defined as identifying a cluster in $\mathcal{C}$-space against the background (q.v. Ting et al. 2012; Ting, Conroy & Goodman 2015). In searching for dissolved clusters, the choice of elements depends on the target. OCs are highly homogeneous so many elements are prospective candidates, while GCs are inhomogeneous in some or all elements, but have distinctive anti-correlations, e.g., O/Mg vs. Na/Al. Martell et al. (2016) use the latter signature to identify stars in the Galactic halo from long-dissolved GCs. Ting, Conroy & Goodman (2015) search for clustering of abundances in the APOGEE survey of the α-rich disk, but do not identify any previously-unknown clusters. Related techniques applied to the α-poor disk using the GALAH survey have met with more success (Quillen et al. 2015; Kos et al. 2018a). Even when clusters are found, however, there may not be a simple one-to-one mapping between $\mathcal{C}$-space clusters and OCs. For example, De Silva et al. (2015) identify a pair of OCs that have identical radial velocities and overlapping stars in $\mathcal{C}$-space. These were likely formed in close proximity, so abundance analysis does correctly identify real structures, but they are not necessarily the same clusters one would identify in physical space.

As for kinematics, strong chemical tagging suffers from a poorly-constrained false positive rate. Ness et al. (2018) identify pairs of field stars with identical abundance patterns that are unlikely to have been born in the same cloud, but the overall frequency of such chance pairs remains uncertain. The problem should be less severe at low [Fe/H], where there are fewer stars and thus a reduced background of interlopers. This led Bland-Hawthorn et al. (2010) to propose applying chemical tagging to the Milky Way's dwarf satellites rather than in its disk. Prima facie, this appears to work (e.g., Karlsson et al. 2012; Webster, Frebel & Bland-Hawthorn 2016), and if confirmed the patterns identified thus far are evidence of star clusters significantly more metal-poor than the lowest metallicity GCs, which have [Fe/H] ≈ -2.4.

The challenges of strong chemical tagging mean that, thus far, it has been most effective when used to characterize groups found by other methods, and to test for potential membership. For example, most stellar streams identified to date appear to be inhomogeneous, implying that they are not disrupting star clusters, and are instead either stars trapped by disc resonances (e.g., Bensby et al. 2007) or the remnants of much more massive ($> 10^8 M_\odot$) systems (e.g., Hasselquist et al. 2017). However, a few streams have proven to be homogeneous, e.g., the 2-Gyr old moving group HR 1614 (Feltzing & Holmberg 2000; De Silva et al. 2007), first identified by Eggen (1978) and confirmed with *Hipparcos* (Eggen 1998). For these systems, elemental homogeneity provides strong evidence that the progenitor is a disrupted star cluster. In the future, it seems likely that the most powerful applications of chemical tagging will continue to be in conjunction with other methods (e.g., Quillen et al. 2015) rather than by itself. Kos et al. (2018a) provide an impressive example of how such combinations can work, by combining chemical tagging and *Gaia* proper motions to identify Pleiades members several degrees from the parent cluster. To date, other identified streams for which chemical tagging may prove useful are faint, so follow-up work awaits multi-object high resolution spectrographs on the ELTs (e.g., MANIFEST on GMT).

6. CONCLUSIONS AND FUTURE PROSPECTS

Star clusters stand at a crossroads of scales. Traditionally star cluster research has focused on clusters as discrete entities whose formation, evolution, and eventual dissolution can be viewed in isolation, at most treating the background galaxy as a source of a passive tidal field. In this paradigm, groups of stars can be neatly classified into open clusters, globular clusters, associations, and field stars, each with distinct properties and formation histories. Such isolation is no longer viable. Observations now reveal that the objects that go on to form bound star clusters are merely the innermost parts of a hierarchy that extends to the scales of galaxies. Once clusters form, the tidal perturbations to which they are subject are affected first by the immediate star-forming environment, and over longer times by the full cosmological history of galaxy assembly. Going to high redshift brings us to a point where the distinction between open clusters and globular clusters dissolves. Consequently any modern understanding of star clusters must bridge from sub-pc to cosmological scales. While many questions remain, the outlines of a model are starting to become clear:

SUMMARY POINTS

1. Stars form in giant molecular clouds (GMCs) over a few free-fall times, at a rate per free-fall time $\epsilon_{ff} \sim 1\%$, with a spread of ~ 0.5 dex. At formation, the stars are hierarchically-structured and cannot be separated neatly into clusters. Distinct star clusters emerge only after gas is cleared, which occurs in less than a few million years. After about 30 Myr in Milky Way-like galaxies, $\sim 10\%$ of the stars are in clusters. There are good theoretical arguments that this fraction should vary with galactic environment, but the evidence either for or against variation is thus far unconvincing.

2. The portions of clouds that form bound clusters have at most mildly elevated ϵ_{ff} values, but form stars over multiple free-fall times, allowing them to reach elevated total star formation efficiencies. Extended star formation renders these regions dynamically relaxed and well-mixed in abundance. In contrast, the stars that are not part of the bound regions never dynamically relax, and may or may not be well-mixed. They only gradually drift apart; they are not part of a coherent flow either collapsing towards or expanding away from the dense regions that become bound.

3. Star formation is ultimately terminated by feedback, but which type of feedback is dominant depends on the local environment. Outflows dominate in clouds too small to contain massive stars,

and radiation pressure dominates in the most massive and dense systems, but for the bulk of star clusters photoionization is likely to be most important. Feedback sets the timescale for gas clearing, and thus determines the density at which star formation transitions from taking place over one to many dynamical times, marking the boundary between the bound and unbound parts of a hierarchical star-forming cloud. The initial cluster mass function (ICMF) of both bound and unbound clusters is well described by a power-law $dN/dM \propto M^{\alpha_M}$ with $\alpha_M \approx -2$, likely with an environmentally-dependent high-mass cutoff. Evidence for any stronger variation in cluster formation with environment is weak.

4. In the first $\approx 10 - 100$ Myr after a gas-free cluster population emerges, it is subject to mass loss via stellar evolution-driven expansion and tidal shocking by gas in its immediate environment. The rate at which the cluster age function (CAF) declines from $\approx 10 - 100$ Myr, as parameterized by its slope α_T $(dN/dT \propto T^{\alpha_T})$, provides a powerful constraint on the importance of these processes. In the Solar neighborhood $\alpha_T \approx -0.5$, implying moderately strong cluster destruction, but the CAF slope outside the Milky Way is uncertain because present samples are compromised by uncertainties in cluster selection and age assignment. The strength of early mass loss mechanisms depends strongly on the distributions of cluster concentration and ratio of size to tidal radius at the end of gas clearing, which are poorly constrained both theoretically and observationally, though there are observational hints that clusters in Milky Way-like galaxies are born fairly concentrated and under-filling their tidal radii. There is tension between these hints and the Milky Way cluster age distribution, which requires moderately strong early cluster disruption.

5. Over timescales of Gyr, the dominant cluster destruction mechanisms are relaxation and tidal shocking (by GMCs for disk clusters, by disk traversal for halo clusters). Relaxation preferentially destroy low-mass clusters; tidal shocking likely does as well, though this depends on how strongly density increases with mass. These mechanisms thus provide a plausible mechanism for transforming the power-law mass distribution ubiquitously observed for young clusters into the peaked distribution observed for globular clusters. However, enhanced tidal shocking during clusters' first ≈ 100 Myr of life also preferentially destroys low-mass clusters, and potentially provides an alternate explanation. The relative importance of the various processes, and the timescale over which the CMF shape changes, likely depends on the galactic environment.

6. Even after clusters come apart in physical space, the stars remain coherent in kinematic, action-angle, and chemical space for timescales of tens, hundreds, and thousands of Myr, respectively. In principle it should be possible to reconstruct clusters in these spaces. However, efforts to do so are still in their infancy, and most present techniques still suffer from high rates of false positives.

FUTURE ISSUES

1. A recurring theme of this review is the uncertainty in inferences of various quantities – cutoffs in the CMF, cluster ages, etc. – based on integrated light. When one cross-checks these results against more accurate methods based on resolved stellar populations and young stellar objects (YSOs), the level of agreement is often worse than a naive interpretation of the stated error bars would suggest. However, integrated light observations will remain a vital tool for the foreseeable future, since even in the era of *JWST* we will not be able to resolve stellar populations in clusters at even ~ 100 Mpc distances, let alone at high redshift. Consequently, there is an urgent need for a program of checking integrated light techniques against resolved stellar populations, and developing new methods that

return realistic error bars.

2. There is a pressing need to extend extragalactic cluster catalogs to lower mass and larger age. At present the Milky Way and extragalactic samples are almost completely non-overlapping in mass, and the extragalactic samples have very limited dynamic range in mass, making it hard to draw conclusions. The best prospects for improving this situation are in pushing to lower mass clusters in galaxies at distances of a few Mpc. This will become possible with extreme AO-fed instrumentation on the next generation of extremely large telescopes. The same instruments will also make it possible to probe star formation in cosmologically denser regimes compared to the low overdensities in the Local Volume.

3. Both analytic models and simulations of star cluster formation that begin from isolated clouds, ignoring the galactic context, are likely reaching the end of their utility. They consistently fail to reproduce observed star formation histories and stellar kinematic, and are likely incapable of correctly predicting the cluster formation efficiency overall. The future is in simulations that start self-consistently from the galactic or cosmological scale, including all relevant physics, and then zoom in to individual clusters, or that do form and evolve clusters using semi-analytic models calibrated to zoom simulations.

4. The largest theoretical uncertainties in cluster demographics now lie at the transition between gas-dominated and gas-free evolution. Pure N-body simulations of the gas-free phase have advanced to the point where it is now possible to simulate a cluster of 10^6 stars directly, but a continuing effort to push these simulations to ever-larger numbers of stars and longer integrations may have hit the point of diminishing returns because the largest uncertainty is the choice of initial conditions, which cannot be addressed without modeling the gas.

5. Observations of 6D phase space from ESA *Gaia* and ESO Gravity may address the issue of the poorly known initial conditions, since they can simultaneously supply initial conditions to the N-body modelers and provide a benchmark against which gas simulators can test their work. This project will, however, require new statistical tools. Statistics like the CMF are problematic for unrelaxed populations that cannot be uniquely decomposed into clusters, while structure-agnostic statistics such as two point correlation functions are less informative than statistics like the CMF when applied to populations of relaxed, discrete clusters. Observed star-forming complexes such as Orion, however, appear to transition smoothly between smooth and relaxed at high density and fractal and unrelaxed at low density. We need statistical tools that can characterize this transition and work across it.

6. The most powerful future methods for reconstructing disrupted clusters are likely to combine elemental abundance and kinematic data. Each method suffers from significant false positives by itself, but the false positives to which each are subject are quite different, and thus the two methods can be used to cross-check one another. However, this will require the availability of abundance measurements going to fainter magnitudes than is now possible. Ground-layer, AO-corrected wide-field spectrographs planned for the next generation of telescopes will allow detailed abundance analysis on stars to fainter limits ($V \approx 20$) than possible today, greatly aiding this effort.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Adamo A, Bastian N. 2018. *The Lifecycle of Clusters in Galaxies*. In *The Birth of Star Clusters*, ed. S Stahler, vol. 424 of *Astrophysics and Space Science Library* 1.1

Adamo A, Kruijssen JMD, Bastian N, Silva-Villa E, Ryon J. 2015. *MNRAS* 452:246–260 2.2, 5, 8

Adamo A, Östlin G, Zackrisson E. 2011. *MNRAS* 417:1904–1912 8, 2.4

Adamo A, Östlin G, Zackrisson E, Hayes M, Cumming RJ, Micheva G. 2010. *MNRAS* 407:870–890 2.1.2

Adamo A, Ryon JE, Messa M, Kim H, Grasha K, et al. 2017. *ApJ* 841:131 4, 2.1.2, 2.2, 5, 6, 2.3, 4.4.1

Adams FC. 2000. *ApJ* 542:964–973 3.5.2

Adams FC, Myers PC. 2001. *ApJ* 553:744–753 3.5.1

Allison RJ, Goodwin SP, Parker RJ, de Grijs R, Portegies Zwart SF, Kouwenhoven MBN. 2009. *ApJ* 700:L99–L103 3.5.2

Andrews JE, Calzetti D, Chandar R, Elmegreen BG, Kennicutt RC, et al. 2014. *ApJ* 793:4 3.3.1

Antoja T, Helmi A, Bienayme O, Bland-Hawthorn J, Famaey B, et al. 2012. *MNRAS* 426:L1–L5 5.1

Antoja T, Helmi A, Romero-Gomez M, Katz D, Babusiaux C, et al. 2018. *ArXiv e-prints* 5.1, 5.2

Armillotta L, Krumholz MR, Fujimoto Y. 2018. *MNRAS* 481:5000–5013 3.6, 3.6

Armus L, Mazzarella JM, Evans AS, Surace JA, Sanders DB, et al. 2009. *PASP* 121:559 1.1

Ascenso J, Alves J, Lago MTVT. 2009. *A&A* 495:147–155 2.1.1

Astropy Collaboration, Robitaille TP, Tollerud EJ, Greenfield P, Droettboom M, et al. 2013. *A&A* 558:A33 1

Azimlu M, Martínez-Galarza JR, Muench AA. 2015. *AJ* 150:95 3.4.1

Ballesteros-Paredes J, Hartmann LW, Vázquez-Semadeni E, Heitsch F, Zamora-Avilés MA. 2011. *MNRAS* 411:65–70 3.2.2

Bally J. 2016. *Annual Review of Astronomy and Astrophysics* 54:491–528 3.3.1

Baraffe I, Chabrier G, Gallardo J. 2009. *ApJ* 702:L27–L31 3.4.1

Baraffe I, Vorobyov E, Chabrier G. 2012. *ApJ* 756:118 3.4.1

Barmby P, McLaughlin DE, Harris WE, Harris GLH, Forbes DA. 2007. *AJ* 133:2764–2786 2, 9

Barnes AT, Longmore SN, Battersby C, Bally J, Kruijssen JMD, et al. 2017. *MNRAS* 469:2263–2285 9

Bastian N. 2008. *MNRAS* 390:759–768 2.2, 2.4

Bastian N, Adamo A, Gieles M, Lamers HJGLM, Larsen SS, et al. 2011. *MNRAS* 417:L6–L10 2.1.2

Bastian N, Adamo A, Gieles M, Silva-Villa E, Lamers HJGLM, et al. 2012. *MNRAS* 419:2606–2622 2.1.2, 2.1.2, 2.2, 4.4.1

Bastian N, Ercolano B, Gieles M, Rosolowsky E, Scheepmaker RA, et al. 2007. *MNRAS* 379:1302–1312 2.7

Bastian N, Lardo C. 2018. *Annual Review of Astronomy and Astrophysics* 56:83–136 1.2.2, 2.3, 2.6

Baumgardt H. 2001. *MNRAS* 325:1323–1331 4.2.2

Baumgardt H, Hilker M. 2018. *MNRAS* 478:1520–1557 2, 9, 4.2.1

Baumgardt H, Makino J. 2003. *MNRAS* 340:227–246 4.2.1, 4.2.2, 4.4.1

Baumgardt H, Parmentier G, Anders P, Grebel EK. 2013. *MNRAS* 430:676–685 5

Baumgardt H, Parmentier G, Gieles M, Vesperini E. 2010. *MNRAS* 401:1832–1838 4.2.2

Beccari G, Petr-Gotzens MG, Boffin HMJ, Romaniello M, Fedele D, et al. 2017. *A&A* 604:A22 3.4.1

Beerman LC, Johnson LC, Fouesneau M, Dalcanton JJ, Weisz DR, et al. 2012. *ApJ* 760:104 2.1.1

Bekki K. 2010. *MNRAS* 401:2753–2762 4.3.2

Bensby T, Feltzing S, Lundström I, Ilyin I. 2005. *A&A* 433:185–203 5.3

Bensby T, Oey MS, Feltzing S, Gustafsson B. 2007. *ApJ* 655:L89–L92 5.1, 5.3

Bertoldi F, McKee CF. 1992. *ApJ* 395:140–157 3.1.1, 3.1.1

Bhattacharya S, Mishra I, Vaidya K, Chen WP. 2017. *ApJ* 847:138 4.2.2

Bianchini P, Renaud F, Gieles M, Varri AL. 2015. *MNRAS* 447:L40–L44 4.4.2

Billett OH, Hunter DA, Elmegreen BG. 2002. *AJ* 123:1454–1475 3.1.2

Binney J, Tremaine S. 2008. *Galactic Dynamics: Second Edition.* Princeton University Press 4.3.1, 4.3.1, 5.2

Birnboim Y, Federrath C, Krumholz M. 2018. *MNRAS* 473:2144–2159 3.2.2

Bisnovatyi-Kogan GS, Silich SA. 1995. *Reviews of Modern Physics* 67:661–712 3.3.5

Bland-Hawthorn J, Freeman KC. 2004. *PASA* 21:110–120 3.6

Bland-Hawthorn J, Gerhard O. 2016. *ARA&A* 54:529–596 19

Bland-Hawthorn J, Karlsson T, Sharma S, Krumholz M, Silk J. 2010. *ApJ* 721:582–596 5.3

Bland-Hawthorn J, Krumholz MR, Freeman K. 2010. *ApJ* 713:166–179 3.6

Boesgaard AM, Lum MG, Deliyannis CP. 2015. *ApJ* 799:202 4.3.2

Bolatto AD, Leroy AK, Rosolowsky E, Walter F, Blitz L. 2008. *ApJ* 686:948–965 3.1.1, 3.1.1

Borissova J, Georgiev L, Hanson MM, Clarke JRA, Kurtev R, et al. 2012. *A&A* 546:A110 3

Borissova J, Ivanov VD, Hanson MM, Georgiev L, Minniti D, et al. 2008. *A&A* 488:151–159 3

Boutloukos SG, Lamers HJGLM. 2003. *MNRAS* 338:717–732 2.3, 4.4.1, 4.4.1, 4.4.1

Bovy J, Bahmanyar A, Fritz TK, Kallivayalil N. 2016. *ApJ* 833:31 5.1

Bovy J, Erkal D, Sanders JL. 2017. *MNRAS* 466:628–668 5.1

Bowden A, Belokurov V, Evans NW. 2015. *MNRAS* 449:1391–1400 5.1

Bragaglia A, Carretta E, D'Orazi V, Sollima A, Donati P, et al. 2017. *A&A* 607:A44 2.6

Bragaglia A, Sneden C, Carretta E, Gratton RG, Lucatello S, et al. 2014. *ApJ* 796:68 2.6

Bressert E, Bastian N, Gutermuth R, Megeath ST, Allen L, et al. 2010. *MNRAS* 409:L54–L58 2.1.1, 2.7

Brockamp M, Küpper AHW, Thies I, Baumgardt H, Kroupa P. 2014. *MNRAS* 441:150–171 4.4.2

Brodie JP, Strader J. 2006. *ARA&A* 44:193–267 2.2, 4.4.2

Brunetti M, Chiappini C, Pfenniger D. 2011. *A&A* 534:A75 4.3.2

Brunt CM, Federrath C, Price DJ. 2010. *MNRAS* 403:1507–1515 3.1.2

Burkhart B. 2018. *ApJ* 863:118 3.2.2

Burkhart B, Mocz P. 2018. *ApJ* Submitted, arXiv:1805.11104 3.2.2

Caldwell S, Chang P. 2018. *MNRAS* 474:4818–4823 3.4.1

Calura F, Few CG, Romano D, D'Ercole A. 2015. *ApJ* 814:L14 3.3.5, 3.3.6

Calzetti D, Armus L, Bohlin RC, Kinney AL, Koornneef J, Storchi-Bergmann T. 2000. *ApJ* 533:682–695 3

Calzetti D, Lee JC, Sabbi E, Adamo A, Smith LJ, et al. 2015. *AJ* 149:51 1.1, 1.2.1

Cantat-Gaudin T, Jordi C, Vallenari A, Bragaglia A, Balaguer-Núñez L, et al. 2018. *ArXiv e-prints* 4.3.1, 5.1

Carlberg RG. 2016. *ApJ* 820:45 5.1

Carlberg RG. 2018. *ApJ* 861:69 4.4.2

Carraro G, Bensby T. 2009. *MNRAS* 397:L106–L109 4.3.2

Carraro G, Villanova S, Demarque P, McSwain MV, Piotto G, Bedin LR. 2006. *ApJ* 643:1151–1159 4.3.2

Carretta E, Bragaglia A, Gratton RG, Recio-Blanco A, Lucatello S, et al. 2010. *A&A* 516:A55 2.6

Chabrier G. 2005. *The Initial Mass Function: from Salpeter 1955 to 2005.* In *The Initial Mass Function 50 Years Later*, eds. E Corbelli, F Palla, H Zinnecker, vol. 327 of *Astrophysics and Space Science Library*. Dordrecht: Springer 3, 3.3.1

Chandar R, Fall SM, Whitmore BC. 2015. *ApJ* 810:1 5

Chandar R, Fall SM, Whitmore BC, Mulia AJ. 2017. *ApJ* 849:128 6, 8, 2.4

Chandar R, Whitmore BC, Calzetti D, Di Nino D, Kennicutt RC, et al. 2011. *ApJ* 727:88 2.1.2

Chandar R, Whitmore BC, Calzetti D, O'Connell R. 2014. *ApJ* 787 2.1.2

Chandar R, Whitmore BC, Dinino D, Kennicutt RC, Chien LH, et al. 2016. *ApJ* 824:71 2.1.2, 5, 2, 9

Chandar R, Whitmore BC, Kim H, Kaleida C, Mutchler M, et al. 2010. *ApJ* 719:966–978 2.1.2

Chandrasekhar S. 1943. *ApJ* 97:255 4.3.2

Chernoff DF, Weinberg MD. 1990. *ApJ* 351:121–156 3.1.1, 4.2.1

Choi J, Dotter A, Conroy C, Cantiello M, Paxton B, Johnson BD. 2016. *ApJ* 823:102 3

Clark JS, Negueruela I, Davies B, Larionov VM, Ritchie BW, et al. 2009. *A&A* 498:109–114 3

Clark PC, Bonnell IA, Zinnecker H, Bate MR. 2005. *MNRAS* 359:809–818 3.5.2

Clarkson WI, Ghez AM, Morris MR, Lu JR, Stolte A, et al. 2012. *ApJ* 751:132 3

Collins DC, Kritsuk AG, Padoan P, Li H, Xu H, et al. 2012. *ApJ* 750:13 3.1.2

Colombo D, Hughes A, Schinnerer E, Meidt SE, Leroy AK, et al. 2014. *ApJ* 784:3 3.1.2

Crocker RM, Krumholz MR, Thompson TA, Baumgardt H, Mackey D. 2018a. *MNRAS* 481:4895–4906 3.3.4, 3.3.7

Crocker RM, Krumholz MR, Thompson TA, Clutterbuck J. 2018b. *MNRAS* 478:81–94 3.3.4

Crutcher RM. 2012. *ARA&A* 50:29–63 3.2.2

Cunningham AJ, Krumholz MR, McKee CF, Klein RI. 2018. *MNRAS* 476:771–792 3.2.2

Da Costa GS, Held EV, Saviane I. 2014. *MNRAS* 438:3507–3520 2.6

Da Rio N, Robberto M, Hillenbrand LA, Henning T, Stassun KG. 2012. *ApJ* 748:14 2.1.1

Da Rio N, Robberto M, Soderblom DR, Panagia N, Hillenbrand LA, et al. 2009. *ApJS* 183:261–277 2

Da Rio N, Tan JC, Covey KR, Cottaar M, Foster JB, et al. 2016. *ApJ* 818:59 3.4.1

Da Rio N, Tan JC, Jaehnig K. 2014. *ApJ* 795:55 3.4.1, 13

da Silva RL, Fumagalli M, Krumholz M. 2012. *ApJ* 745:145 1, 3, 3.3.1

Dalcanton JJ, Williams BF, Lang D, Lauer TR, Kalirai JS, et al. 2012. *ApJS* 200:18 1.1, 1.2.1

Dale JE. 2017. *MNRAS* 467:1067–1082 3.3.2

Dale JE, Ercolano B, Bonnell IA. 2012. *MNRAS* 424:377–392 3.3.2

Dale JE, Ercolano B, Bonnell IA. 2013. *MNRAS* 430:234–246 3.3.2

Davies B, Bastian N, Gieles M, Seth AC, Mengel S, Konstantopoulos IS. 2011. *MNRAS* 411:1386–1394 3.4.2

Davies B, Figer DF, Kudritzki RP, MacKenty J, Najarro F, Herrero A. 2007. *ApJ* 671:781–801 3

Davis SW, Jiang YF, Stone JM, Murray N. 2014. *ApJ* 796:107 3.3.4

de Meulenaer P, Narbutis D, Mineikis T, Vansevičius V. 2013. *A&A* 550:A20 2.1.2

De Silva GM, Freeman KC, Bland-Hawthorn J, Asplund M, Bessell MS. 2007. *AJ* 133:694–704 5.3

De Silva GM, Freeman KC, Bland-Hawthorn J, Asplund M, Williams M, Holmberg J. 2011. *MNRAS* 415:563–575 2.6

De Silva GM, Freeman KC, Bland-Hawthorn J, Martell S, de Boer EW, et al. 2015. *MNRAS* 449:2604–2617 1.1, 3.6, 5.3

De Silva GM, Sneden C, Paulson DB, Asplund M, Bland-Hawthorn J, et al. 2006. *AJ* 131:455–460 2.6

De Simone R, Wu X, Tremaine S. 2004. *MNRAS* 350:627–643 5.1

de Zeeuw PT, Hoogerwerf R, de Bruijne JHJ, Brown AGA, Blaauw A. 1999. *AJ* 117:354–399 5.1

Dehnen W. 1998. *AJ* 115:2384–2396 5.1

Dehnen W. 2000. *AJ* 119:800–812 4.3.2

Dobbs CL, Adamo A, Few CG, Calzetti D, Dale DA, et al. 2017. *MNRAS* 464:3580–3596 3.5.3

Dobbs CL, Burkert A, Pringle JE. 2011. *MNRAS* 413:2935–2942 3.2.2

Dobbs CL, Krumholz MR, Ballesteros-Paredes J, Bolatto AD, Fukui Y, et al. 2014. *Protostars and Planets VI* :3–26 14

Dotter A, Conroy C, Cargile P, Asplund M. 2017. *ApJ* 840:99 2.6

Dunne BC, Chu YH, Chen CHR, Lowry JD, Townsley L, et al. 2003. *ApJ* 590:306–313 3.3.5

Eggen OJ. 1978. *ApJ* 222:191–208 5.3

Eggen OJ. 1998. *AJ* 115:2453–2458 5.3

Elmegreen BG. 2000. *ApJ* 530:277–281 2.7

Elmegreen BG. 2006. *ApJ* 648:572–579 3.5.1, 3.5.2

Elmegreen BG. 2008. *ApJ* 672:1006–1012 3.5.2

Elmegreen BG. 2010. *ApJ* 712:L184–L188 4.3.1

Elmegreen BG, Efremov YN. 1997. *ApJ* 480:235–245 1.2.2

Elmegreen BG, Elmegreen DM. 2001. *AJ* 121:1507–1511 2.7

Elmegreen BG, Falgarone E. 1996. *ApJ* 471:816 3.5.1.1

Elmegreen BG, Hunter DA. 2010. *ApJ* 712:604–623 4.3.1, 4.4.1

Elson RA, Fall SM. 1988. *AJ* 96:1383 2.1.2, 2.4

Elson RAW, Fall SM. 1985. *ApJ* 299:211–218 2.1.2, 2.4

Elson RAW, Fall SM, Freeman KC. 1987. *ApJ* 323:54 1, 4.1

Elson RAW, Freeman KC, Lauer TR. 1989. *ApJ* 347:L69 4.2.1

Emerick A, Bryan GL, Mac Low MM. 2019. *MNRAS* 482:1304–1329 3.6

Escala A, Larson RB. 2008. *ApJ* 685:L31 3.1.2, 3.1.2

Evans II NJ, Heiderman A, Vutisalchavakul N. 2014. *ApJ* 782:114 3.2.1, 10

Faesi CM, Lada CJ, Forbrich J. 2018. *ApJ* 857:19 3.1.1, 3.1.2

Falgarone E, Pety J, Hily-Blant P. 2009. *A&A* 507:355–368 3.1.1

Fall SM, Chandar R. 2012. *ApJ* 752:96 5, 4.2.1, 4.4.1

Fall SM, Chandar R, Whitmore BC. 2005. *ApJ* 631:L133–L136 4.4.1

Fall SM, Chandar R, Whitmore BC. 2009. *ApJ* 704:453–468 4.4.1, 4.4.1, 4.4.1

Fall SM, Krumholz MR, Matzner CD. 2010. *ApJ* 710:L142–L146 3.3, 3.3.3, 3.3.6, 3.5.1.1, 3.5.1.2

Fall SM, Zhang Q. 2001. *ApJ* 561:751–765 2.2, 4.2.2, 4.4.1

Faucher-Giguère CA, Kaspi VM. 2006. *ApJ* 643:332–355 4.2.1

Federrath C. 2013. *MNRAS* 436:3167–3172 3.2.1

Federrath C. 2015. *MNRAS* 450:4035–4042 3.2.2

Federrath C, Klessen RS. 2012. *ApJ* 761:156 3.2.2

Feltzing S, Holmberg J. 2000. *A&A* 357:153–163 5.3

Feng Y, Krumholz MR. 2014. *Nature* 513:523–525 2.6, 3.6

Fierlinger KM, Burkert A, Ntormousi E, Fierlinger P, Schartmann M, et al. 2016. *MNRAS* 456:710–730 3.3.5

Figer DF, MacKenty JW, Robberto M, Smith K, Najarro F, et al. 2006. *ApJ* 643:1166–1179 3

Figer DF, McLean IS, Morris M. 1999. *ApJ* 514:202–220 3

Fleck Jr. RC. 1996. *ApJ* 458:739 3.5.1.1

Forbes DA, Bastian N, Gieles M, Crain RA, Kruijssen JMD, et al. 2018. *Proceedings of the Royal Society of London Series A* 474:20170616 1.1, 4.4.2

Fouesneau M, Johnson LC, Weisz DR, Dalcanton JJ, Bell EF, et al. 2014. *ApJ* 786:117 6, 2.4, 2, 9

Fouesneau M, Lançon A. 2010. *A&A* 521:A22 2.1.2

Fouesneau M, Lançon A, Chandar R, Whitmore BC. 2012. *ApJ* 750:60 2.1.2

Freeman K, Bland-Hawthorn J. 2002. *ARA&A* 40:487–537 3.6, 5.3

Friel ED. 1995. *Annual Review of Astronomy and Astrophysics* 33:381–414 4.3.2

Fukushige T, Heggie DC. 1995. *MNRAS* 276:206–218 4.2.1

Fukushige T, Heggie DC. 2000. *MNRAS* 318:753–761 4.2.2

Gallagher MJ, Leroy AK, Bigiel F, Cormier D, Jiménez-Donaire MJ, et al. 2018. *ApJ* 858:90 3.2.1, 10

Gao Y, Solomon PM. 2004. *ApJ* 606:271–290 10

García-Burillo S, Usero A, Alonso-Herrero A, Graciá-Carpio J, Pereira-Santaella M, et al. 2012. *A&A* 539:A8 3.2.1

Gatto A, Walch S, Naab T, Girichidis P, Wünsch R, et al. 2017. *MNRAS* 466:1903–1924 3.3.5

Gavagnin E, Bleuler A, Rosdahl J, Teyssier R. 2017. *MNRAS* 472:4155–4172 3.3.2

Gazak JZ, Davies B, Bastian N, Kudritzki R, Bergemann M, et al. 2014. *ApJ* 787:142 2.4

Geen S, Hennebelle P, Tremblin P, Rosdahl J. 2015. *MNRAS* 454:4484–4502 3.3.2

Geen S, Hennebelle P, Tremblin P, Rosdahl J. 2016. *MNRAS* 463:3129–3142 3.3.2, 3.3.6

Geisler D, Villanova S, Carraro G, Pilachowski C, Cummings J, et al. 2012. *ApJ* 756:L40 2.6

Gennaro M, Brandner W, Stolte A, Henning T. 2011. *MNRAS* 412:2469–2488 3

Gentry ES, Krumholz MR, Dekel A, Madau P. 2017. *MNRAS* 465:2471–2488 3.3.6

Gentry ES, Krumholz MR, Madau P, Lupi A. 2018. *MNRAS* Submitted, arXiv:1802.06860 3.3.5, 3.3.6

Genzel R, Newman S, Jones T, Förster Schreiber NM, Shapiro K, et al. 2011. *ApJ* 733:101 3.1.2

Gerhard O. 2001. *ApJ* 546:L39–L42 4.3.2

Getman KV, Feigelson ED, Kuhn MA. 2014. *ApJ* 787:109 3.4.1, 3.4.2

Getman KV, Feigelson ED, Kuhn MA, Bate MR, Broos PS, Garmire GP. 2018. *MNRAS* 476:1213–1223 2.1.1, 3.4.1, 3.4.2, 3.4.2

Gieles M. 2009. *MNRAS* 394:2113–2126 4.4.1

Gieles M, Athanassoula E, Portegies Zwart SF. 2007. *MNRAS* 376:809–819 4.3.1, 4.3.2

Gieles M, Baumgardt H. 2008. *MNRAS* 389:L28–L32 4.2.2

Gieles M, Baumgardt H, Heggie DC, Lamers HJGLM. 2010. *MNRAS* 408:L16–L20 4.2.1

Gieles M, Heggie DC, Zhao H. 2011. *MNRAS* 413:2509–2524 4.2.2

Gieles M, Portegies Zwart SF. 2011. *MNRAS* 410:L6–L7 1.2.1

Gieles M, Portegies Zwart SF, Baumgardt H, Athanassoula E, Lamers HJGLM, et al. 2006. *MNRAS* 371:793–804 4.3.1, 4.3.1, 4.3.1

Gieles M, Renaud F. 2016. *MNRAS* 463:L103–L107 4.3.1

Ginsburg A, Bressert E, Bally J, Battersby C. 2012. *ApJ* 758:L29 3.4.2

Ginsburg A, Kruijssen JMD. 2018. *ApJ* 864:L17 8, 2.4

Gnedin OY, Ostriker JP. 1997. *ApJ* 474:223–255 4.3.1

Gnedin OY, Ostriker JP. 1999. *ApJ* 513:626–637 4.3.1

Goddard QE, Bastian N, Kennicutt RC. 2010. *MNRAS* 405:857–869 8, 2.4, 2.4

Goldbaum NJ, Krumholz MR, Matzner CD, McKee CF. 2011. *ApJ* 738:101–+ 3.2.2, 3.4.2

González-Fernández C, Negueruela I. 2012. *A&A* 539:A100 3

Goudfrooij P, Fall SM. 2016. *ApJ* 833:8 4.2.2

Gouliermis DA. 2018. *PASP* 130:072001 1.2.2, 2.7, 3.6

Gouliermis DA, Elmegreen BG, Elmegreen DM, Calzetti D, Cignoni M, et al. 2017. *MNRAS* 468:509–530 2.7

Grasha K, Calzetti D, Adamo A, Kim H, Elmegreen BG, et al. 2017a. *ApJ* 840:113 2.7, 4.3.1

Grasha K, Elmegreen BG, Calzetti D, Adamo A, Aloisi A, et al. 2017b. *ApJ* 842:25 2.7

Gratton RG, Carretta E, Bragaglia A. 2012. *Asron. Astrop. Rev.* 20:50 2.6

Griffin RF. 1998. *The Observatory* 118:223–225 5.1

Grillmair CJ, Johnson R. 2006. *ApJ* 639:L17–L20 4.2.2, 5.1

Grudić MY, Hopkins PF, Faucher-Giguère CA, Quataert E, Murray N, Kereš D. 2018. *MNRAS* 475:3511–3528 3.2.2, 3.3.3, 3.3.7, 3.4.2, 14

Guszejnov D, Hopkins PF, Grudić MY. 2018. *MNRAS* 477:5139–5149 3.5.1.1

Haid S, Walch S, Seifried D, Wünsch R, Dinnbier F, Naab T. 2018. *MNRAS* 478:4799–4815 3.3.5

Harayama Y, Eisenhauer F, Martins F. 2008. *ApJ* 675:1319–1342 3

Harper-Clark E, Murray N. 2009. *ApJ* 693:1696–1712 3.3.5

Harris WE, Pudritz RE. 1994. *ApJ* 429:177–191 3.1.2

Hasselquist S, Shetrone M, Smith V, Holtzman J, McWilliam A, et al. 2017. *ApJ* 845:162 5.3

Hawkins K, Jofré P, Masseron T, Gilmore G. 2015. *MNRAS* 453:758–774 5.3

Heiderman A, Evans II NJ, Allen LE, Huard T, Heyer M. 2010. *ApJ* 723:1019–1037 3.2.1

Heithausen A, Bensch F, Stutzki J, Falgarone E, Panis JF. 1998. *A&A* 331:L65–L68 3.1.2

Heitsch F, Ballesteros-Paredes J, Hartmann L. 2009. *ApJ* 704:1735–1742 3.2.2

Helmi A. 2008. *Asron. Astrop. Rev.* 15:145–188 5.1, 5.2

Helmi A, Babusiaux C, Koppelman HH, Massari D, Veljanoski J, Brown AGA. 2018a. *Nature* 563:85–88 5.2

Helmi A, Navarro JF, Nordström B, Holmberg J, Abadi MG, Steinmetz M. 2006. *MNRAS* 365:1309–1323 5.1

Helmi A, van Leeuwen F, McMillan PJ, Massari D, Antoja T, et al. 2018b. *A&A* 616:A12 4.3.2

Helmi A, Veljanoski J, Breddels MA, Tian H, Sales LV. 2017. *A&A* 598:A58 5.1

Hénault-Brunet V, Gieles M, Evans CJ, Sana H, Bastian N, et al. 2012. *A&A* 545:L1 3.4.2

Hennebelle P, Chabrier G. 2008. *ApJ* 684:395–410 3.5.1.1

Hennebelle P, Chabrier G. 2011. *ApJ* 743:L29 3.2.2

Hennekemper E, Gouliermis DA, Henning T, Brandner W, Dolphin AE. 2008. *ApJ* 672:914–929 2.7

Hénon M. 1961. *Annales d'Astrophysique* 24:369. English translation by F. Renaud available as arXiv:1103.3499

4.2.2

Hénon M. 1965a. *Annales d'Astrophysique* 28:499 4.2.2

Hénon M. 1965b. *Annales d'Astrophysique* 28:992 4.2.2

Heyer M, Gutermuth R, Urquhart JS, Csengeri T, Wienen M, et al. 2016. *A&A* 588:A29 3.2.1, 10

Heyer M, Krawczyk C, Duval J, Jackson JM. 2009. *ApJ* 699:1092–1103 3.1.1

Heyer MH, Carpenter JM, Snell RL. 2001. *ApJ* 551:852–866 3.1.1

Hillenbrand LA, Hartmann LW. 1998. *ApJ* 492:540–553 3.4.1

Hills JG. 1980. *ApJ* 235:986–991 3.5.2

Hollyhead K, Adamo A, Bastian N, Gieles M, Ryon JE. 2016. *MNRAS* 460:2087–2102 5

Hollyhead K, Bastian N, Adamo A, Silva-Villa E, Dale J, et al. 2015. *MNRAS* 449:1106–1117 3.3.6

Hong J, Kim E, Lee HM, Spurzem R. 2013. *MNRAS* 430:2960–2972 4.2.2

Hopkins PF. 2013. *MNRAS* 430:1653–1693 3.2.2

Hopkins PF. 2014. *ApJ* 797:59 3.6

Hopkins PF, Lee H. 2016. *MNRAS* 456:4174–4190 3.6

Hopkins PF, Murray N, Quataert E, Thompson TA. 2010. *MNRAS* 401:L19–L23 2.5, 3.3.7

Hopkins PF, Narayanan D, Murray N, Quataert E. 2013. *MNRAS* 433:69–77 3.2.1

Hopkins PF, Wetzel A, Kereš D, Faucher-Giguère CA, Quataert E, et al. 2018. *MNRAS* 480:800–863 3.3.4

Hosokawa T, Offner SSR, Krumholz MR. 2011. *ApJ* 738:140 3.4.1

Howard CS, Pudritz RE, Harris WE. 2017. *MNRAS* 470:3346–3358 3.3.2

Howard CS, Pudritz RE, Harris WE. 2018. *Nature Astronomy* 2:725–730 3.3.3

Hughes A, Meidt SE, Colombo D, Schinnerer E, Pety J, et al. 2013. *ApJ* 779:46 3.1.1

Hunter JD. 2007. *Computing in Science & Engineering* 9:90–95 1

Immer K, Schuller F, Omont A, Menten KM. 2012. *A&A* 537:A121 14

Innanen KA, Harris WE, Webbink RF. 1983. *AJ* 88:338–360 4.2.2

Jackson JM, Rathborne JM, Shah RY, Simon R, Bania TM, et al. 2006. *ApJS* 163:145–159 1.1

Jacquet E, Krumholz MR. 2011. *ApJ* 730:116–+ 3.3.4

Jaehnig KO, Da Rio N, Tan JC. 2015. *ApJ* 798:126 3.4.1

Jameson KE, Bolatto AD, Leroy AK, Meixner M, Roman-Duval J, et al. 2016. *ApJ* 825:12 11, 11

Jensen SS, Haugbølle T. 2018. *MNRAS* 474:1176–1193 3.4.1

Jethwa P, Torrealba G, Navarrete C, Carballo-Bello JA, de Boer T, et al. 2018. *MNRAS* 480:5342–5351 5.1

Jílková L, Carraro G, Jungwiert B, Minchev I. 2012. *A&A* 541:A64 4.3.2

Jog CJ, Solomon PM. 1992. *ApJ* 387:152–161 3.1.2

Johnson KE, Brogan CL, Indebetouw R, Testi L, Wilner DJ, et al. 2018. *ApJ* 853:125 3.1.2

Johnson KE, Leroy AK, Indebetouw R, Brogan CL, Whitmore BC, et al. 2015a. *ApJ* 806:35 3.1.2

Johnson LC, Seth AC, Dalcanton JJ, Beerman LC, Fouesneau M, et al. 2016. *ApJ* 827:33 8, 2.4

Johnson LC, Seth AC, Dalcanton JJ, Beerman LC, Fouesneau M, et al. 2017. *ApJ* 839:78 5, 2.2

Johnson LC, Seth AC, Dalcanton JJ, Caldwell N, Fouesneau M, et al. 2012. *ApJ* 752:95 2, 9

Johnson LC, Seth AC, Dalcanton JJ, Wallace ML, Simpson RJ, et al. 2015b. *ApJ* 802:127 2.1.1

Jordán A, McLaughlin DE, Côté P, Ferrarese L, Peng EW, et al. 2007. *ApJS* 171:101–145 5, 2.2, 2.2, 5

Kamann S, Bastian N, Husser TO, Martocchia S, Usher C, et al. 2018. *MNRAS* 480:1689–1695 3.4.2

Karlsson T, Bland-Hawthorn J, Freeman KC, Silk J. 2012. *ApJ* 759:111 5.3

Katz D, Sartoretti P, Cropper M, Panuzzo P, Seabroke GM, et al. 2018. *A&A* Submitted, arXiv:1804.09372 5.2

Kauffmann J, Goldsmith PF, Melnick G, Tolls V, Guzman A, Menten KM. 2017. *A&A* 605:L5 10

Kauffmann J, Pillai T, Goldsmith PF. 2013. *ApJ* 779:185 3.1.2, 3.2.2

Kawamura A, Mizuno Y, Minamidani T, Filipović MD, Staveley-Smith L, et al. 2009. *ApJS* 184:1–17 2.7

Kepley AA, Morrison HL, Helmi A, Kinman TD, Van Duyne J, et al. 2007. *AJ* 134:1579–1595 5.1

Keto ER, Myers PC. 1986. *ApJ* 304:466–480 3.1.1

Kharchenko NV, Piskunov AE, Röser S, Schilbach E, Scholz RD. 2005. *A&A* 440:403–408 4.4.1

Kharchenko NV, Piskunov AE, Schilbach E, Röser S, Scholz RD. 2013. *A&A* 558:A53 1.2.2, 2, 9, 4.2.1

Kim D, Lu JR, Konopacky Q, Chu L, Toller E, et al. 2019. *ApJ* :submitted, arXiv:1812.04134 3.4.1, 3.4.2

Kim JG, Kim WT, Ostriker EC. 2018. *ApJ* 859:68 3.3.2, 3.3.3, 3.4.2, 3.5.1.2

Kim Jh, Ma X, Grudić MY, Hopkins PF, Hayward CC, et al. 2018. *MNRAS* 474:4232–4244 3.5.3, 4.4.2

Kim WT, Ostriker EC. 2001. *ApJ* 559:70–95 3.1.2, 3.1.2

Kimm T, Cen R, Rosdahl J, Yi SK. 2016. *ApJ* 823:52 3.5.3, 4.4.2

King IR. 1962. *AJ* 67:471 4.1

King IR. 1966. *AJ* 71:64 1, 4.1

Klessen RS, Hennebelle P. 2010. *A&A* 520:A17+ 3.2.2, 3.4.2

Koo BC, McKee CF. 1992. *ApJ* 388:103 3.3.5

Koposov SE, Rix HW, Hogg DW. 2010. *ApJ* 712:260–273 5.1

Kordopatis G, Binney J, Gilmore G, Wyse RFG, Belokurov V, et al. 2015. *MNRAS* 447:3526–3535 4.3.2

Körtgen B, Seifried D, Banerjee R, Vázquez-Semadeni E, Zamora-Avilés M. 2016. *MNRAS* 459:3460–3474 3.3.6

Kos J, Bland-Hawthorn J, Freeman K, Buder S, Traven G, et al. 2018a. *MNRAS* 473:4612–4633 5.3

Kos J, de Silva G, Buder S, Bland-Hawthorn J, Sharma S, et al. 2018b. *MNRAS* 480:5242–5259 1.1

Kounkel M, Covey K, Suárez G, Román-Zúñiga C, Hernandez J, et al. 2018. *AJ* 156:84 2, 13, 3.4.2

Kraus AL, Hillenbrand LA. 2008. *ApJ* 686:L111–L114 2.7

Krause MGH, Charbonnel C, Bastian N, Diehl R. 2016. *A&A* 587:A53 3.3.5

Kreckel K, Faesi C, Kruijssen JMD, Schruba A, Groves B, et al. 2018. *ApJ* 863:L21 3.2.1

Kroupa P, Aarseth S, Hurley J. 2001. *MNRAS* 321:699–712 3.4.2, 3.5.2

Kruijssen JMD. 2009. *A&A* 507:1409–1423 4.2.1

Kruijssen JMD. 2012. *MNRAS* 426:3008–3040 2.4, 3.3.6, 3.4.2, 3.5.2, 3.5.2, 4.3.1, 4.4.2

Kruijssen JMD. 2014. *Classical and Quantum Gravity* 31:244006 1.1, 3.1.2

Kruijssen JMD. 2015. *MNRAS* 454:1658–1686 4.3.1

Kruijssen JMD, Bastian N. 2016. *MNRAS* 457:L24–L28 2.4

Kruijssen JMD, Maschberger T, Moeckel N, Clarke CJ, Bastian N, Bonnell IA. 2012a. *MNRAS* 419:841–853 3.5.2

Kruijssen JMD, Pelupessy FI, Lamers HJGLM, Portegies Zwart SF, Bastian N, Icke V. 2012b. *MNRAS* 421:1927–1941 4.4.1, 4.4.2

Kruijssen JMD, Pelupessy FI, Lamers HJGLM, Portegies Zwart SF, Icke V. 2011. *MNRAS* 414:1339–1364 2.3, 4.3.1, 4.3.1, 4.4.1, 4.4.2

Kruijssen JMD, Schruba A, Hygate APS, Hu CY, Haydon DT, Longmore SN. 2018. *MNRAS* 479:1866–1952 3.2.1, 3.3

Krumholz MR. 2014. *Phys. Rep.* 539:49–134 3.2, 10

Krumholz MR. 2018. *MNRAS* 480:3468–3482 3.3.3

Krumholz MR, Adamo A, Fumagalli M, Calzetti D. 2018. *MNRAS* In press, arXiv:1810.10173 2.1.2, 4.4.1

Krumholz MR, Adamo A, Fumagalli M, Wofford A, Calzetti D, et al. 2015a. *ApJ* 812:147 2.1.2, 2.1.2, 2.4

Krumholz MR, Bate MR, Arce HG, Dale JE, Gutermuth R, et al. 2014. *Protostars and Planets VI* :243–266 1.1

Krumholz MR, Dekel A, McKee CF. 2012. *ApJ* 745:69 3.2.1

Krumholz MR, Fumagalli M, da Silva RL, Rendahl T, Parra J. 2015b. *MNRAS* 452:1447–1467 1, 3, 3.3.1

Krumholz MR, Klein RI, McKee CF. 2012. *ApJ* 754:71 3.3.1, 12, 3.4.2

Krumholz MR, Matzner CD. 2009. *ApJ* 703:1352–1362 3.3.3

Krumholz MR, McKee CF. 2005. *ApJ* 630:250–268 3.1.2, 3.2.2

Krumholz MR, Tan JC. 2007. *ApJ* 654:304–315 3.2.1

Krumholz MR, Thompson TA. 2012. *ApJ* 760:155 3.3.4

Krumholz MR, Ting YS. 2018. *MNRAS* 475:2236–2252 3.6

Kuhn MA, Feigelson ED, Getman KV, Baddeley AJ, Broos PS, et al. 2014. *ApJ* 787:107 2, 1.2.1

Kuhn MA, Hillenbrand LA, Sills A, Feigelson ED, Getman KV. 2018. *ApJ* In press, arXiv:1807.02115 3.4.2, 3.4.2

Kundic T, Ostriker JP. 1995. *ApJ* 438:702–707 4.3.1

Kuznetsova A, Hartmann L, Ballesteros-Paredes J. 2015. *ApJ* 815:27 3.4.2

Kuznetsova A, Hartmann L, Ballesteros-Paredes J. 2018. *MNRAS* 473:2372–2377 3.4.2

Lada CJ, Lada EA. 2003. *ARA&A* 41:57–115 1.2.1, 1.2.2, 2.1.1, 3.3.7, 3.5.1, 3.5.2

Lada CJ, Lombardi M, Roman-Zuniga C, Forbrich J, Alves JF. 2013. *ApJ* 778:133 3.1.1, 3.2.1, 10

Lada CJ, Margulis M, Dearborn D. 1984. *ApJ* 285:141–152 3.5.2, 15

Lamb JB, Oey MS, Werk JK, Ingleby LD. 2010. *ApJ* 725:1886–1902 2.2

Lamers HJGLM, Baumgardt H, Gieles M. 2010. *MNRAS* 409:305–328 4.2.1, 4.2.2, 4.4.1, 4.4.1

Lamers HJGLM, Gieles M, Bastian N, Baumgardt H, Kharchenko NV, Portegies Zwart S. 2005. *A&A* 441:117–129 4.2.2, 4.4.1, 4.4.1, 4.4.1, 4.4.1

Larsen SS. 2002. *AJ* 124:1393–1409 2.2

Larsen SS. 2009. *A&A* 494:539–551 2.2

Larsen SS, de Mink SE, Eldridge JJ, Langer N, Bastian N, et al. 2011. *A&A* 532:A147 2.1.1

Larsen SS, Richtler T. 2000. *A&A* 354:836–846 2.4

Larson RB. 1981. *MNRAS* 194:809–826 3.1.1

Lee EJ, Miville-Deschênes MA, Murray NW. 2016. *ApJ* 833:229 3.2.1, 10, 3.2.1

Lee H, Hopkins PF, Squire J. 2017. *MNRAS* 469:3532–3540 3.6

Lee YN, Hennebelle P. 2016a. *A&A* 591:A30 3.2.2, 3.4.2

Lee YN, Hennebelle P. 2016b. *A&A* 591:A31 3.2.2, 3.4.2

Leitherer C, Schaerer D, Goldader JD, Delgado RMG, Robert C, et al. 1999. *ApJS* 123:3–40 3

Leroy AK, Bolatto AD, Ostriker EC, Walter F, Gorski M, et al. 2018. *ApJ* In press, arXiv:1804.02083 2, 9, 3.1.2, 3.1.2

Leroy AK, Schinnerer E, Hughes A, Kruijssen JMD, Meidt S, et al. 2017. *ApJ* 846:71 3.2.1, 10

Li H, Gnedin OY, Gnedin NY. 2018. *ApJ* 861:107 3.5.3, 4.4.2

Li H, Gnedin OY, Gnedin NY, Meng X, Semenov VA, Kravtsov AV. 2017. *ApJ* 834:69 3.5.3, 4.4.2

Li H, Li D, Qian L, Xu D, Goldsmith PF, et al. 2015. *ApJS* 219:20 3.3.1

Lim S, Lee MG. 2015. *ApJ* 804:123 8

Linden ST, Evans AS, Rich J, Larson KL, Armus L, et al. 2017. *ApJ* 843:91 5, 6

Liu F, Asplund M, Yong D, Meléndez J, Ramírez I, et al. 2016a. *MNRAS* 463:696–704 2.6

Liu F, Yong D, Asplund M, Ramírez I, Meléndez J. 2016b. *MNRAS* 457:3934–3948 2.6, 3.6

Lombardi M, Alves J, Lada CJ. 2015. *A&A* 576:L1 3.1.2

Longmore SN, Kruijssen JMD, Bastian N, Bally J, Rathborne J, et al. 2014. *Protostars and Planets VI* :291–314 3.3.6, 3.4.2, 3.4.2

Longmore SN, Rathborne J, Bastian N, Alves J, Ascenso J, et al. 2012. *ApJ* 746:117 14

Lopez LA, Krumholz MR, Bolatto AD, Prochaska JX, Ramirez-Ruiz E. 2011. *ApJ* 731:91 3.3.5

Lopez LA, Krumholz MR, Bolatto AD, Prochaska JX, Ramirez-Ruiz E, Castro D. 2014. *ApJ* 795:121 3.3.5

Lynden-Bell D, Kalnajs AJ. 1972. *MNRAS* 157:1 4.3.2

Mackey AD, Gilmore GF. 2003a. *MNRAS* 338:120–130 4.1

Mackey AD, Gilmore GF. 2003b. *MNRAS* 338:85–119 4.1

Mackey J, Gvaramadze VV, Mohamed S, Langer N. 2015. *A&A* 573:A10 3.3.5

Madrid JP, Leigh NWC, Hurley JR, Giersz M. 2017. *MNRAS* 470:1729–1737 4.2.2, 4.3.1

Maíz Apellániz J. 2009. *ApJ* 699:1938–1955 2.1.2

Maji M, Zhu Q, Li Y, Charlton J, Hernquist L, Knebe A. 2017. *ApJ* 844:108 3.5.3

Mamikonyan EN, McMillan SLW, Vesperini E, Mac Low MM. 2017. *ApJ* 837:70 4.4.2

Mandelker N, van Dokkum PG, Brodie JP, van den Bosch FC, Ceverino D. 2018. *ApJ* 861:148 4.4.2

Marino AF, Yong D, Milone AP, Piotto G, Lundquist M, et al. 2018. *ApJ* 859:81 2.6

Martell SL, Shetrone MD, Lucatello S, Schiavon RP, Mészáros S, et al. 2016. *ApJ* 825:146 5.3

Martinez-Medina LA, Pichardo B, Peimbert A, Moreno E. 2017. *ApJ* 834:58 4.3.1

Mathieu RD. 1983. *ApJ* 267:L97–L101 3.5.2

Matzner CD. 2002. *ApJ* 566:302–314 3.3, 3.3.2, 3.3.6

Matzner CD, Jumper PH. 2015. *ApJ* 815:68 3.3, 3.3.1, 3.4.2

Matzner CD, McKee CF. 2000. *ApJ* 545:364–378 3.3.1

McCrady N, Graham JR. 2007. *ApJ* 663:844–856 1.1, 5, 2, 9

McKee CF, Ostriker EC. 2007. *ARA&A* 45:565–687 3.1.2, 3.2

McKee CF, Parravano A, Hollenbach DJ. 2015. *ApJ* 814:13 1.2.1

McKee CF, Tan JC. 2003. *ApJ* 585:850–871 3.1.1

McKee CF, van Buren D, Lazareff B. 1984. *ApJ* 278:L115–L118 3.3.6

McKee CF, Williams JP. 1997. *ApJ* 476:144–+ 2.2

McLaughlin DE. 2000. *ApJ* 539:618–640 4.2.1

McLaughlin DE, Fall SM. 2008. *ApJ* 679:1272–1287 4.2.2

McLaughlin DE, van der Marel RP. 2005. *ApJS* 161:304–360 4.2.1

McMillan PJ. 2011. *MNRAS* 414:2446–2457 5.2

Meidt SE, Leroy AK, Rosolowsky E, Kruijssen JMD, Schinnerer E, et al. 2018. *ApJ* 854:100 3.2.2

Meléndez J, Bedell M, Bean JL, Ramírez I, Asplund M, et al. 2017. *A&A* 597:A34 2.6

Messa M, Adamo A, Calzetti D, Reina-Campos M, Colombo D, et al. 2018a. *MNRAS* 477:1683–1707 5, 2.2, 6, 2.3

Messa M, Adamo A, Östlin G, Calzetti D, Grasha K, et al. 2018b. *MNRAS* 473:996–1018 2.1.2, 5, 6, 2.3, 8, 4.4.1

Messina S, Parihar P, Distefano E. 2017. *MNRAS* 468:931–945 3.4.1

Miholics M, Kruijssen JMD, Sills A. 2017. *MNRAS* 470:1421–1435 4.3.1, 4.4.1

Miholics M, Webb JJ, Sills A. 2014. *MNRAS* 445:2872–2877 4.4.2

Miholics M, Webb JJ, Sills A. 2016. *MNRAS* 456:240–247 4.4.2

Millman K, Aivazis M. 2011. *Computing in Science & Engineering* 13:9–12 1

Minchev I, Famaey B. 2010. *ApJ* 722:112–121 4.3.2

Minchev I, Famaey B, Combes F, Di Matteo P, Mouhcine M, Wozniak H. 2011. *A&A* 527:A147 4.3.2

Miville-Deschênes MA, Murray N, Lee EJ. 2017. *ApJ* 834:57 3.1.1, 3.1.2

Mok A, Chandar R, Fall SM. 2018. *ApJ* In press, arXiv:1806.11192 2.2

Motte F, Bontemps S, Louvet F. 2018. *Annual Review of Astronomy and Astrophysics* 56:41–82 3.4.2

Mulia AJ, Chandar R, Whitmore BC. 2016. *ApJ* 826:32 5

Murray D, Goyal S, Chang P. 2018. *MNRAS* 475:1023–1035 3.3.1, 12

Murray N, Chang P. 2015. *ApJ* 804:44 3.2.2

Murray N, Quataert E, Thompson TA. 2010. *ApJ* 709:191–209 3.3.3, 3.3.4

Murray N, Rahman M. 2010. *ApJ* 709:424–435 3.3.2

Murray SD, Lin DNC. 1990. *ApJ* 357:105–112 3.6

Myers AT, Klein RI, Krumholz MR, McKee CF. 2014. *MNRAS* 439:3420–3438 3.2.2

Naiman JP, Ramirez-Ruiz E, Lin DNC. 2018. *MNRAS* 478:2794–2811 3.3.5

Nakamura F, Li ZY. 2014. *ApJ* 783:115 3.3.1

Negueruela I, González-Fernández C, Marco A, Clark JS. 2011. *A&A* 528:A59 3

Negueruela I, González-Fernández C, Marco A, Clark JS, Martínez-Núñez S. 2010. *A&A* 513:A74 3

Ness M, Rix HW, Hogg DW, Casey AR, Holtzman J, et al. 2018. *ApJ* 853:198 5.3

Nieva MF, Przybilla N. 2012. *A&A* 539:A143 4.3.2

Ochsendorf BB, Meixner M, Chastenet J, Tielens AGGM, Roman-Duval J. 2016. *ApJ* 832:43 11, 3.2.1

Ochsendorf BB, Meixner M, Roman-Duval J, Rahman M, Evans II NJ. 2017. *ApJ* 841:109 3.2.1, 10, 3.2.1, 11, 11

Odenkirchen M, Brosche P, Geffert M, Tucholke HJ. 1997. *New Astron.* 2:477–499 4.2.2, 4.3.1

Odenkirchen M, Grebel EK, Dehnen W, Rix HW, Cudworth KM. 2002. *AJ* 124:1497–1510 4.2.2, 4.3.1

Offner SSR, Arce HG. 2015. *ApJ* 811:146 3.3.5

Offner SSR, Chaban J. 2017. *ApJ* 847:104 3.3.1

Oliphant T. 2007. *Computing in Science & Engineering* 9:10–20 1

Onus A, Krumholz MR, Federrath C. 2018. *MNRAS* In press, arXiv:1801.09952 3.2.1, 10, 10

Origlia L, Valenti E, Rich RM, Ferraro FR. 2006. *ApJ* 646:499–504 4.3.2

Ostriker JP, Spitzer Jr. L, Chevalier RA. 1972. *ApJ* 176:L51 4.3.1

Padoan P, Cambrésy L, Juvela M, Kritsuk A, Langer WD, Norman ML. 2006. *ApJ* 649:807–815 3.6

Padoan P, Federrath C, Chabrier G, Evans II NJ, Johnstone D, et al. 2014. *Protostars and Planets VI* :77–100 3.1.1

Padoan P, Haugbølle T, Nordlund Å. 2012. *ApJ* 759:L27 3.2.2

Padoan P, Nordlund Å. 2011. *ApJ* 730:40–+ 3.2.2

Palla F, Stahler SW. 2000. *ApJ* 540:255–270 3.4.1

Parmentier G, Pfalzner S. 2013. *A&A* 549:A132 3.5.2

Pellegrini EW, Baldwin JA, Ferland GJ. 2011. *ApJ* 738:34 3.3.5

Pellerin A, Meyer M, Harris J, Calzetti D. 2007. *ApJ* 658:L87–L90 2.4

Pfeffer J, Kruijssen JMD, Crain RA, Bastian N. 2018. *MNRAS* 475:4309–4346 4.4.2

Piffl T, Binney J, McMillan PJ, Steinmetz M, Helmi A, et al. 2014. *MNRAS* 445:3133–3151 5.2

Piskunov AE, Just A, Kharchenko NV, Berczik P, Scholz RD, et al. 2018. *A&A* 614:A22 2.3, 6, 7, 7

Piskunov AE, Schilbach E, Kharchenko NV, Röser S, Scholz RD. 2007. *A&A* 468:151–161 9

Plunkett AL, Arce HG, Corder SA, Dunham MM, Garay G, Mardones D. 2015. *ApJ* 803:22 3.3.1

Popescu B, Hanson MM, Elmegreen BG. 2012. *ApJ* 751:122 2.1.2

Portegies Zwart SF, McMillan SLW, Gieles M. 2010. *ARA&A* 48:431–493 1.2.1, 1.2.1, 2.5, 2.5, 8, 4

Price-Whelan AM, Bonaca A. 2018. *ApJ* 863:L20 5.1

Prieto JL, Gnedin OY. 2008. *ApJ* 689:919–935 4.3.1, 4.4.2

Quillen AC, Anguiano B, De Silva G, Freeman K, Zucker DB, et al. 2015. *MNRAS* 450:2354–2366 5.3

Quillen AC, De Silva G, Sharma S, Hayden M, Freeman K, et al. 2018a. *MNRAS* 5.1

Quillen AC, Minchev I, Bland-Hawthorn J, Haywood M. 2009. *MNRAS* 397:1599–1606 4.3.2

Quillen AC, Nolting E, Minchev I, De Silva G, Chiappini C. 2018b. *MNRAS* 475:4450–4466 4.3.2

Rahner D, Pellegrini EW, Glover SCO, Klessen RS. 2017. *MNRAS* 470:4453–4472 3.3, 3.3.5

Raskutti S, Ostriker EC, Skinner MA. 2016. *ApJ* 829:130 3.3.3

Reggiani M, Robberto M, Da Rio N, Meyer MR, Soderblom DR, Ricci L. 2011. *A&A* 534:A83 2.1.1, 3.4.1

Reina-Campos M, Kruijssen JMD. 2017. *MNRAS* 469:1282–1298 3.1.2, 4.4.2

Reissl S, Klessen RS, Mac Low MM, Pellegrini EW. 2018. *A&A* 611:A70 3.3.4

Renaud F. 2018. *New Astron. Rev.* 81:1–38 1.1, 4

Renaud F, Agertz O, Gieles M. 2017. *MNRAS* 465:3622–3636 4.4.2

Renaud F, Bournaud F, Duc PA. 2015. *MNRAS* 446:2038–2054 3.5.3

Renaud F, Gieles M. 2013. *MNRAS* 431:L83–L87 4.4.2

Renaud F, Gieles M, Boily CM. 2011. *MNRAS* 418:759–769 4.1, 18

Renzini A. 2013. *Mem. Soc. Astron. Ital.* 84:162 1.2.2, 2.6

Rey-Raposo R, Dobbs C, Agertz O, Alig C. 2017. *MNRAS* 464:3536–3551 3.3.6

Rice TS, Goodman AA, Bergin EA, Beaumont C, Dame TM. 2016. *ApJ* 822:52 3.1.1, 3.1.2

Ricotti M, Parry OH, Gnedin NY. 2016. *ApJ* 831:204 3.5.3, 4.4.2

Riedel AR, Blunt SC, Lambrides EL, Rice EL, Cruz KL, Faherty JK. 2017. *AJ* 153:95 5.1

Rieder S, Ishiyama T, Langelaan P, Makino J, McMillan SLW, Portegies Zwart S. 2013. *MNRAS* 436:3695–3706 4.4.2

Robberto M, Soderblom DR, Bergeron E, Kozhurina-Platais V, Makidon RB, et al. 2013. *ApJS* 207:10 1

Robertson B, Goldreich P. 2012. *ApJ* 750:L31 3.2.2

Rochau B, Brandner W, Stolte A, Gennaro M, Gouliermis D, et al. 2010. *ApJ* 716:L90–L94 3

Rogers H, Pittard JM. 2013. *MNRAS* 431:1337–1351 3.3.5, 3.3.6

Rogers H, Pittard JM. 2014. *MNRAS* 441:964–982 3.3.5

Roman-Duval J, Jackson JM, Heyer M, Rathborne J, Simon R. 2010. *ApJ* 723:492–507 3.1.1

Rosdahl J, Teyssier R. 2015. *MNRAS* 449:4380–4403 3.3.4

Rosen AL, Lopez LA, Krumholz MR, Ramirez-Ruiz E. 2014. *MNRAS* 442:2701–2716 3.3.5

Rossi LJ, Hurley JR. 2015a. *MNRAS* 454:1453–1467 4.4.2

Rossi LJ, Hurley JR. 2015b. *MNRAS* 446:3389–3403 4.4.2

Roškar R, Debattista VP, Quinn TR, Stinson GS, Wadsley J. 2008. *ApJ* 684:L79–L82 4.3.2

Ryon JE, Adamo A, Bastian N, Smith LJ, Gallagher III JS, et al. 2014. *AJ* 148:33 5, 6, 8

Ryon JE, Bastian N, Adamo A, Konstantopoulos IS, Gallagher JS, et al. 2015. *MNRAS* 452:525–539 2, 9, 4.2.1, 4.2.2, 4.4.1

Ryon JE, Gallagher JS, Smith LJ, Adamo A, Calzetti D, et al. 2017. *ApJ* 841:92 2, 9, 4.2.1, 4.2.2, 4.4.1

Salim DM, Federrath C, Kewley LJ. 2015. *ApJ* 806:L36 3.2.1

Schmeja S. 2011. *Astronomische Nachrichten* 332:172 1.2.1

Schmeja S, Gouliermis DA, Klessen RS. 2009. *ApJ* 694:367–375 2.7

Schneider N, Ossenkopf V, Csengeri T, Klessen RS, Federrath C, et al. 2015. *A&A* 575:A79 3.1.1, 3.1.2, 3.5.2

Schuller F, Menten KM, Contreras Y, Wyrowski F, Schilke P, et al. 2009. *A&A* 504:415–427 1.1

Sellwood JA. 2010. *MNRAS* 409:145–155 5.2

Sellwood JA, Binney JJ. 2002. *MNRAS* 336:785–796 4.3.2

Shukirgaliyev B, Parmentier G, Berczik P, Just A. 2017. *A&A* 605:A119 3.5.2

Silich S, Tenorio-Tagle G. 2018. *MNRAS* 478:5112–5122 3.3.5

Silva-Villa E, Adamo A, Bastian N, Fouesneau M, Zackrisson E. 2014. *MNRAS* 440:L116–L120 6, 2.3

Silva-Villa E, Larsen SS. 2011. *A&A* 529:A25 2.1.1

Skinner MA, Ostriker EC. 2015. *ApJ* 809:187 3.3.4

Smith R, Fellhauer M, Goodwin S, Assmann P. 2011. *MNRAS* 414:3036–3043 3.5.2

Smith R, Goodwin S, Fellhauer M, Assmann P. 2013. *MNRAS* 428:1303–1311 3.5.2

Soderblom DR, Hillenbrand LA, Jeffries RD, Mamajek EE, Naylor T. 2014. *Ages of Young Stars*. In *Protostars and Planets VI* 3.4.1

Sokal KR, Johnson KE, Indebetouw R, Massey P. 2016. *ApJ* 826:194 3.3.6

Sollima A, Martínez-Delgado D, Valls-Gabaud D, Peñarrubia J. 2011. *ApJ* 726:47 4.2.2

Solomon PM, Rivolo AR, Barrett J, Yahil A. 1987. *ApJ* 319:730–741 3.1.1, 3.1.1

Solway M, Sellwood JA, Schönrich R. 2012. *MNRAS* 422:1363–1383 5.3

Souto D, Cunha K, Smith VV, Allende Prieto C, García-Hernández DA, et al. 2018. *ApJ* 857:14 2.6

Spitzer Lyman J. 1958. *ApJ* 127:17 4.3.1

Steinmetz M, Zwitter T, Siebert A, Watson FG, Freeman KC, et al. 2006. *AJ* 132:1645–1668 1.1

Stephens IW, Jackson JM, Whitaker JS, Contreras Y, Guzmán AE, et al. 2016. *ApJ* 824:29 10

Sun J, Leroy AK, Schruba A, Rosolowsky E, Hughes A, et al. 2018. *ApJ* 860:172 3.1.1, 3.5.1.2

Sun W, de Grijs R, Fan Z, Cameron E. 2016. *ApJ* 816:9 2.2

Sur S, Pan L, Scannapieco E. 2014. *ApJ* 784:94 3.6

Takahashi K, Portegies Zwart SF. 2000. *ApJ* 535:759–775 4.2.1

Tan JC, Kong S, Butler MJ, Caselli P, Fontani F. 2013. *ApJ* 779:96 3.1.2

Tan JC, Krumholz MR, McKee CF. 2006. *ApJ* 641:L121–L124 3.4.1

Tanikawa A, Fukushige T. 2010. *PASJ* 62:1215–1230 4.2.2

Taylor BJ. 2000. *A&A* 362:563–579 5.1

Thompson TA, Fabian AC, Quataert E, Murray N. 2015. *MNRAS* 449:147–161 3.3.4

Thompson TA, Krumholz MR. 2016. *MNRAS* 455:334–342 3.3.3

Thompson TA, Quataert E, Murray N. 2005. *ApJ* 630:167–185 3.3.4

Ting YS, Conroy C, Goodman A. 2015. *ApJ* 807:104 5.3

Ting YS, Freeman KC, Kobayashi C, De Silva GM, Bland-Hawthorn J. 2012. *MNRAS* 421:1231–1255 5.3

Toalá JA, Arthur SJ. 2018. *MNRAS* 478:1218–1230 3.3.5

Townsley LK, Broos PS, Chu YH, Gruendl RA, Oey MS, Pittard JM. 2011. *ApJS* 194:16 3.3.5

Townsley LK, Broos PS, Feigelson ED, Brandl BR, Chu YH, et al. 2006. *AJ* 131:2140–2163 3.3.5

Townsley LK, Feigelson ED, Montmerle T, Broos PS, Chu YH, Garmire GP. 2003. *ApJ* 593:874–905 3.3.5

Traficante A, Fuller GA, Smith RJ, Billot N, Duarte-Cabral A, et al. 2018. *MNRAS* 473:4975–4985 3.2.2

Tricco TS, Price DJ, Laibe G. 2017. *MNRAS* 471:L52–L56 3.6

Trick WH, Coronado J, Rix HW. 2018. *MNRAS* Submitted, arXiv:1805.03653 5.2

Trumpler RJ. 1930. *Lick Observatory Bulletin* 14:154–188 1.2.1, 1.2.2

Tsang BTH, Milosavljević M. 2015. *MNRAS* 453:1108–1120 3.3.4

Tsang BTH, Milosavljević M. 2018. *MNRAS* 478:4142–4161 3.3.4

Urquhart JS, König C, Giannetti A, Leurini S, Moore TJT, et al. 2018. *MNRAS* 473:1059–1102 3.1.2, 3.4.2, 14, 3.5.2

Usero A, Leroy AK, Walter F, Schruba A, García-Burillo S, et al. 2015. *AJ* 150:115 3.2.1, 10

Utomo D, Sun J, Leroy AK, Kruijssen JMD, Schinnerer E, et al. 2018. *ApJ* 861:L18 3.2.1, 10

Vasiliev E. 2018. *MNRAS* Submitted, arXiv:1807.09775 4.3.2

Vázquez-Semadeni E, González-Samaniego A, Colín P. 2017. *MNRAS* 467:1313–1328 3.4.2

Vesperini E, McMillan SLW, Portegies Zwart S. 2009. *ApJ* 698:615–622 4.2.1

Villanova S, Geisler D, Carraro G, Moni Bidin C, Muñoz C. 2013. *ApJ* 778:186 2.6

Vink JS, de Koter A, Lamers HJGLM. 2001. *A&A* 369:574–588 3.3.5

Vorobyov EI, Elbakyan V, Hosokawa T, Sakurai Y, Guedel M, Yorke H. 2017. *A&A* 605:A77 3.4.1

Vutisalchavakul N, Evans II NJ, Heyer M. 2016. *ApJ* 831:73 3.2.1, 10, 3.2.1

Walker DL, Longmore SN, Bastian N, Kruijssen JMD, Rathborne JM, et al. 2016. *MNRAS* 457:4536–4545 3.4.2

Wang P, Li Z, Abel T, Nakamura F. 2010. *ApJ* 709:27–41 3.2.2, 3.3.1, 3.4.2

Ward JL, Kruijssen JMD. 2018. *MNRAS* 475:5659–5676 3.4.2

Wareing CJ, Pittard JM, Falle SAEG. 2017. *MNRAS* 465:2757–2783 3.3.5, 3.3.6

Webb JJ, Sills A, Harris WE, Hurley JR. 2014. *MNRAS* 445:1048–1055 4.3.1

Webster D, Frebel A, Bland-Hawthorn J. 2016. *ApJ* 818:80 5.3

Whitmore BC, Brogan C, Chandar R, Evans A, Hibbard J, et al. 2014. *ApJ* 795:156 3.1.2

Whitmore BC, Chandar R, Fall SM. 2007. *AJ* 133:1067–1084 4.4.1, 4.4.1

Whitmore BC, Schweizer F, Leitherer C, Borne K, Robert C. 1993. *AJ* 106:1354 2.1.2

Whitmore BC, Zhang Q, Leitherer C, Fall SM, Schweizer F, Miller BW. 1999. *AJ* 118:1551–1576 2.1.2

Whitworth A. 1979. *MNRAS* 186:59–67 3.5.1.2

Williams JP, McKee CF. 1997. *ApJ* 476:166–183 3.3.2

Wilson CD, Scoville N, Madden SC, Charmandaris V. 2000. *ApJ* 542:120–127 3.1.2

Wilson CD, Scoville N, Madden SC, Charmandaris V. 2003. *ApJ* 599:1049–1066 3.1.2

Wolfire MG, Cassinelli JP. 1987. *ApJ* 319:850–867 3.3.4

Wong T, Hughes A, Ott J, Muller E, Pineda JL, et al. 2011. *ApJS* 197:16 11, 11

Woods DT, Klein RI, Castor JI, McKee CF, Bell JB. 1996. *ApJ* 461:767 3.3.2

Wu J, Evans II NJ, Shirley YL, Knez C. 2010. *ApJS* 188:313–357 10

Xu Y, Newberg HJ, Carlin JL, Liu C, Deng L, et al. 2015. *ApJ* 801:105 5.2

Yang AY, Thompson MA, Urquhart JS, Tian WW. 2018. *ApJS* 235:3 3.3

Yeh SCC, Matzner CD. 2012. *ApJ* 757:108 3.3.5

Yong D, Meléndez J, Grundahl F, Roederer IU, Norris JE, et al. 2013. *MNRAS* 434:3542–3565 2.6

Zamora-Avilés M, Vázquez-Semadeni E. 2014. *ApJ* 793:84 3.4.2

Zamora-Avilés M, Vázquez-Semadeni E, Colín P. 2012. *ApJ* 751:77 3.2.2

Zeidler P, Nota A, Grebel EK, Sabbi E, Pasquali A, et al. 2017. *AJ* 153:122 3

Zhang Q, Fall SM. 1999. *ApJ* 527:L81–L84 1.1, 2.1.2

Zinn R. 1985. *ApJ* 293:424–444 1.2.2

Zinnecker H. 2010. *Embedded and open star clusters: basic questions*. In *Star Clusters: Basic Galactic Building Blocks Throughout Time and Space*, eds. R de Grijs, JRD Lépine, vol. 266 of *IAU Symposium* 2.7