THE FORMATION HISTORY OF GLOBULAR CLUSTERS

Dean E. McLaughlin

1Department of Astronomy, 601 Campbell Hall, University of California, Berkeley, CA 94720
2Hubble Fellow

ABSTRACT

The properties of old globular cluster systems in galaxy halos are used to infer quantitative constraints on aspects of star formation that are arguably as relevant in a present-day context as they were during the protogalactic epoch. First, the spatial distribution of globulars in three large galaxies, together with trends in total cluster population vs. galaxy luminosity for 97 early-type systems plus the halo of the Milky Way, imply that bound stellar clusters formed with an essentially universal efficiency throughout early protogalaxies: by mass, always $r_{cl} = 0.26\% \pm 0.05\%$ of star-forming gas was converted into globulars rather than halo field stars. That this fraction is so robust in the face of extreme variations in local and global galaxy environment suggests that any parcel of gas needs primarily to exceed a relative density threshold in order to form a bound cluster of stars. Second, it is shown that a strict scaling between total binding energy, luminosity, and Galactocentric position, $E_b = 7.2 \times 10^{39} \text{erg} \left(\frac{L}{L_\odot}\right)^{2.05} \left(\frac{r_{gc}}{8 \text{kpc}}\right)^{-0.4}$, is a defining equation for a fundamental plane of Galactic globular clusters. The characteristics of this plane, which subsumes all other observable correlations between the structural parameters of globulars, provide a small but complete set of facts that must be explained by theories of cluster formation and evolution in the Milky Way. It is suggested that the $E_b(L, r_{gc})$ relation specifically resulted from star formation efficiencies having been systematically higher inside more massive protoglobular gas clumps.

Key words: galaxies: fundamental parameters – galaxies: star clusters – globular clusters: general – stars: formation large-scale star-formation processes in the early universe. But at the same time, most new stars today, whether in the Galactic disk or in galaxy mergers and starbursts, are born not in isolation but in groups. To be sure, the formation of a bona fide star cluster is a rare event (only a very small fraction of young stellar groups emerge from their natal clumps of gas as gravitationally bound units) but, as will be argued here, it is one that occurs regularly. Thus, individual globular clusters must also be viewed—even if in a limit—as the products of a robust mode of smaller-scale star formation that has always been viable.

It should come as no surprise, then, that there is no definitive theoretical description of the formation history of globular clusters; while some concepts have been identified that do seem likely to survive as elements of a correct theory in the future, at this point there is simply no model that can claim completeness. Detailed discussions of the many theories already in the literature may be found in, e.g., Ashman & Zepf (1998) or Meylan & Heggie (1997). The focus here will instead be on recent progress in extracting quantitative and (as nearly as possible) model-independent constraints from the data on many GCSs. A similarly empirical discussion is given by Harris (2000) with an eye mostly to implications for large-scale aspects of galaxy formation and evolution (see also Ashman & Zepf 1998). In what follows, particular emphasis will be placed on applications to the problem of generic star formation on subgalactic scales.

A serious concern, when trying to use GCSs in this way, is the influence of dynamical evolution on the gross properties of cluster systems that have been immersed for a Hubble time in the tidal fields of their parent galaxies. Two-body relaxation and evaporation, disk- and bulge-shock heating, chaotic scattering or disruption by a compact nucleus, and dynamical friction: all of these processes whittle away at clusters individually and collectively (Spitzer 1987; Aguilar et al. 1988; Ostriker et al. 1989; Capuzzo-Dolcetta 1993; Murali & Weinberg 1997a; Murali & Weinberg 1997b; Murali & Weinberg 1997c; Vesperini 1997). In the case of the Milky Way particularly, the net effect is to define a roughly triangular region in mass–radius space, within which globulars are predicted to survive a 10-Gyr dynamical onslaught (Fall & Rees 1977) — the implication being that any clusters born outside such a “survival triangle” would have disappeared, taking with
them vital information on their birth properties. However, further investigation shows that Milky Way clusters located at large Galactocentric radii \( r_{gc} \gtrsim 3 \) kpc, roughly the effective radius of the bulge, do not fill their expected survival triangles \( \text{(Caputo \& Castellani 1984, Gnedin \& Ostriker 1997)} \). Similarly, evolutionary calculations geared to conditions in the giant elliptical M87 \( \text{(Murai \& Weinberg 1971)} \) suggest that the damage done to an initial GCS may be largely confined to within an effective radius in that galaxy as well. Thus, there are some features of globular clusters, and of GCSs, that are not dictated purely by evolution from some completely obscured initial conditions, and while care must be taken in identifying such observables, the task is not an impossible one.

Following the suggestion that large fractions of the GCSs in ellipticals may have formed in major mergers \( \text{(e.g., Schweizer 1987, Ashman \& Zepf 1992)} \), and the related discovery of young, massive star clusters in systems like the Antennae galaxies \( \text{(Whitmore \& Schweizer 1995)} \), much recent discussion in this field has centered on the interpretation of (often) bi- and multimodal distributions of globular cluster colors (as indicators of integrated metallicity) in relation to their host galaxies’ dynamical histories \( \text{(e.g., Zepf \& Ashman 1993, Forbes et al. 1997, Côté et al. 1998, Kissler-Patig et al. 1998, Kissler-Patig \& Gebhardt 1999, Côté et al. 2000)} \). However, as will be discussed further in §3, metallicity is completely decoupled from the other basic properties of individual globulars in the Milky Way; thus, while GCS colors are of interest in the context of galaxy formation and chemical evolution, they seem to be of marginal importance to the star-formation process itself. In this connection, it is worth noting explicitly that literally thousands of genuinely old globulars have now been identified, in many galaxies besides our own, with colors as red as or redder than solar: high metallicity has never posed any apparent obstacle to the formation of massive star clusters. This cautions against theories of globular formation such as those of Fall \& Rees (1985) and Murray \& Lin (1993), which, at least in their current form, are able to account only for metal-poor objects.

Of rather more direct relevance to the small-scale problem are the luminosity or mass functions of cluster systems. It is well known that the overall range of Galactic globular cluster masses, \( m \sim 10^4 \sim 10^6 \) M\(_\odot\), and the mean value \( \langle m \rangle = 2.4 \times 10^5 \) M\(_\odot\), are essentially universal properties of other GCSs. Traditionally, theories of globular cluster formation attempted to explain just this basic mass scale \( \text{(e.g., Peebles \& Dicke 1968, Fall \& Rees 1985)} \). More recently, however, attention has turned to the full mass spectrum, \( dN/dm \) (the number of clusters with mass \( m \) in a single galaxy), which—in the regime usually observed, \( m \gtrsim 10^5 \) M\(_\odot\)—is similar (though not identical) from galaxy to galaxy; shows no detectable variations with radius in any one system; and is remarkably similar to the mass function, \( dN/dm \propto m^{-1.7} \) or so, of the dense gas clumps currently forming stars in the giant molecular clouds of the Milky Way disk \( \text{(Harris \& Pudritz 1994)} \). These facts have suggested a general physical picture, developed by Harris \& Pudritz (1994), building on earlier arguments by Larson (1988) and Larson (1993), in which handfuls of globulars form in each of many protogalactic fragments whose gas masses (of order \( 10^9 \) M\(_\odot\)), sizes, and internal velocity dispersions correspond to disk GMCs scaled up by some three orders of magnitude in mass. Such a picture has the obvious appeal of being at least conceptually consistent both with current models of hierarchical galaxy formation \( \text{(there is clearly some affinity with the classic picture of Searle \& Zinn 1978 as well)} \) and with the observed pattern of present-day star formation. It also has more specific attractions \( \text{(e.g., the protogalactic fragments and their protoglobulars are presumed to be supported largely by nonthermal mechanisms, thus allowing for metal-rich protoglobulars with the correct mass scale)} \) but many details remain to be worked out. See McLaughlin \& Pudritz (1996) \( \text{(and compare the rather different view of Elmegreen \& Efremov 1997)} \) for an attempt at a quantitative theory for GCS mass spectra which, with the Harris \& Pudritz (1994) framework as a backdrop, makes explicit use of the emerging links between present-day and protogalactic star formation.

In addition to color and luminosity distributions, photometric studies of extragalactic GCSs yield estimates of their specific frequencies (total cluster populations, normalized to the parent galaxy luminosity) and spatial structures (number density of globulars as a function of galactocentric position). These are the focus of §3, where it is shown that a large sample of early-type galaxies display a common ratio of GCS mass to total mass in stars and (hot) gas—an apparently universal efficiency for the formation of globular clusters \( \text{(McLaughlin 1999)} \). Finally, §3 discusses the binding energies of globular clusters in the Milky Way \( \text{(McLaughlin 2000)} \). It is shown that a strong correlation between binding energy, total cluster luminosity, and Galactocentric position is instrumental in defining a fundamental plane for Galactic globulars—which, by incorporating every one of their many other structural correlations, systematically reduces these data to the smallest possible set of independent physical constraints for theories of cluster formation and evolution.

### 2. The Efficiency of Cluster Formation

The possibility that globular cluster systems can be connected not only to galaxy formation, but to ongoing star formation as well, is suggested by the fact that this latter process operates largely in a clustered mode. One dramatic example of this is the situation in the Orion B (L1630) molecular cloud, where 96% of a complete sample of young stellar objects are physically associated with just four dense clumps of gas, each containing \( > 300 \) M\(_\odot\) of material \( \text{(Lada et al. 1991, Lada 1992)} \). More generally, Patec \& Pudritz (1994) note that this is just a result of the dif-
ferent power-law slopes in the mass function of molecular clumps (as above, \(dN/dm \propto m^{-1.7}\)), so that the largest clumps, which always weigh in at \(10^2-10^3 M_\odot\), contain most of the star-forming gas mass in any molecular cloud and the stellar initial mass function (\(dN/dm \propto m^{-2.35}\) putting most of the mass in young stars into \(\lesssim 1 M_\odot\) objects). In more extreme environments, the “super” star clusters—luminous, blue, compact associations with integrated properties roughly consistent with those expected of young globulars—found in many merging and starburst galaxies can account for as much as \(~20\%\) of the UV light from such systems (Meurer et al. 1995).

Again, however, this is not to say that all, or even most, stars are born into true clusters that exist for any length of time as systems with negative energy. At some point during the collapse and fragmentation of a cluster-sized cloud of gas, the massive stars which it has formed will expel any remaining gas by the combined action of their stellar winds, photoionization, and supernova explosions. If the cumulative star formation efficiency (SFE) of the cloud, \(\eta \equiv M_{\text{stars}}/(M_{\text{stars}} + M_{\text{gas}})\), is below a critical threshold when the gas is lost, then the blow-out removes sufficient energy that the stellar group left behind is unbound and disperses into the field. The precise value of this threshold depends on details of the dynamics and magnetic field in the gas cloud before its self-destruction, and on the timescale over which the massive stars disperse the gas; various estimates place it in the range \(\eta_{\text{crit}} \sim 0.2-0.5\) (e.g., Hills 1980; Mathieu 1983; Elmegreen & Clemens 1985). For a discussion of globulars specifically, see Goodwin 1997.

A general theory of star formation must therefore be able to anticipate the final cumulative SFE in any single piece of gas with (say) a given mass and density, and thereby predict whether or not it will form a bound cluster. No such theory yet exists. It is possible, however, to empirically estimate the probability that a cluster-sized cloud of gas is able to achieve \(\eta > \eta_{\text{crit}}\). This probability—or, equivalently, that fraction of an ensemble of massive star-forming clouds which manages to produce bound stellar systems—is referred to here as the efficiency of cluster formation. To get a handle on this for globulars in particular, McLaughlin (1999) works in terms of the mass fraction of

\[
\epsilon_{\text{cl}} \equiv \frac{M_{\text{gcs}}^{\text{init}}}{M_{\text{gas}}^{\text{init}}},
\]

(1)

where \(M_{\text{gcs}}^{\text{init}}\) refers to the total gas supply that was available to form stars in a protogalaxy—whether in a monolithic collapse or a slower assembly of many distinct, subgalactic clumps is unimportant—and \(M_{\text{gas}}^{\text{init}}\) is the total mass of all globulars formed in that gas. The advantage of this definition for \(\epsilon_{\text{cl}}\) is that the total mass of a GCS is expected to be very well preserved over the course of dynamical evolution in a galaxy; presently observed values are reasonable indicators of the initial quantities. This is ultimately due to the fact that GCS mass spectra are shallow enough that (again, like the clumps in Galactic molecular clouds) most of any one system’s mass is contained in its most massive clusters. But most of the destruction processes mentioned in \(\epsilon_{\text{cl}}\) operate most effectively against low-mass globulars, which may be lost in great numbers (substantially affecting the total GCS population, \(N_{\text{tot}}\)) while decreasing the integrated \(M_{\text{gcs}}\) by as little as \(~25\%\) (McLaughlin 1999).

Until very recently, it was generally assumed that a galaxy’s total luminosity, or stellar mass, was an adequate stand-in for \(M_{\text{gas}}^{\text{init}}\). Thus, the number of globulars per unit of halo light has long been taken as a direct tracer of the efficiency of globular cluster formation in galaxies. However, this approach leads quickly to two interesting problems.

2.1. GLOBAL AND LOCAL SPECIFIC FREQUENCIES

Specific frequency was originally defined by Harris & van den Bergh (1981) as a global property of galaxies. It is nominally the ratio, modulo a convenient normalization, of the total GCS population to the total V-band light integrated over an entire galaxy:

\[
S_N \equiv \frac{N_{\text{tot}}}{\langle V \rangle} = \frac{8.55 \times 10^7 \langle N_{\text{tot}}/L_{V, \text{gal}} \rangle}{\langle V \rangle} (2)
\]

Most subsequent studies of GCSs have therefore estimated their total populations and cited \(S_N\)-values according to equation (2). It is more useful, however, following the discussion just above, to refer to total GCS and stellar masses; thus,

\[
S_N \approx 2500 \left( \frac{\langle m \rangle}{2.4 \times 10^5 M_\odot} \right)^{-1} \left( \frac{\Upsilon_{V, \text{gal}}}{7 M_\odot L_{\odot}^{-1}} \right) \frac{M_{\text{gcs}}}{M_{\text{stars}}} \tag{3}
\]

for a standard mean globular cluster mass \((m)\) and a representative stellar mass-to-light ratio, \(\Upsilon_{V, \text{gal}}\), appropriate to the cores of large ellipticals (which value is used so as not to include any nonbaryonic dark matter in the galaxy mass). Note that \((m)\) is not observed to deviate significantly from the Milky Way value, either from galaxy to galaxy or from place to place within any one system, but that \(\Upsilon_{V, \text{gal}}\) does vary systematically, as a function of luminosity, among large ellipticals (e.g., van der Marel 1991).

Global specific frequencies in a large sample of early-type galaxies are shown in Fig. 1. (The GCSs of spirals are generally less populous, and often more difficult to identify, than those in elliptical systems; thus, the data on late-type galaxies are relatively sparse.) The square points correspond to dwarf ellipticals and spheroidals, some in the Local Group (the two faintest objects are the Fornax and Sagittarius dwarfs) and others in the Virgo cluster (Durrell et al. 1996; Miller et al. 1998); filled circles represent regular giant ellipticals in a wide range of field and cluster environments (see, e.g., Harris 1991 and Kissler-Patig 1997); and open circles stand for the centrally located galaxies (which are often also the brightest) in a large number of groups and clusters (Blakeslee et al. 1997; Harris et al. 1998).
Three points are immediately apparent. First, among normal E’s an average $S_N \approx 5$ (roughly, $M_{gcs}/M_{stars} \sim 0.002$) is indicated. Second, the specific frequencies of central galaxies in groups and clusters show a systematic departure from this “typical” value, increasing strongly towards brighter galaxy magnitudes. And third, while the brightest of the dwarf ellipticals have $S_N$ comparable to the giants, the ratio increases towards fainter luminosities in these small systems. All in all, $S_N$ ranges over more than a factor of 20 in early-type galaxies. If the ratio $M_{gcs}/M_{stars}$ were a good approximation to $\epsilon_{cl}$ in equation (1), the implication would be that the basic efficiency of cluster formation also varied drastically—and in a non-monotonic fashion—from galaxy to galaxy. This is the first specific frequency problem. Although it has been much discussed in the literature (see McLaughlin 1999, Elmegreen 2000, or Harris 2000 for recent reviews and references), no satisfactory explanation (or prediction) of it has ever been advanced.

Very closely related to this, local specific frequencies may be defined at different projected radii within a single galaxy, by taking the ratio of its GCS surface (number) density profile, $N_{cl}(R_{gc})$, and its $V$-band light intensity, $I_V(R_{gc})$, normalized as in equation (2). (This is then proportional to the ratio of cluster and field-star mass densities, just as in eq. (1).) Beyond an effective radius or so (where the effects of dynamical evolution on the GCS are presumably minimized), it is found in some galaxies that local specific frequencies increase outwards; in others, they remain constant. Equivalently, some galaxies’ GCSs are significantly more extended than their stellar halos, while others’ GCS and field-star distributions trace each other accurately on large spatial scales. If the simple assumption $S_N \propto \epsilon_{cl}$ were applied here, it would appear to suggest that bound stellar clusters were sometimes (but not always and for reasons completely unknown) more likely to form at larger distances from the centers of galaxies, in gas that was presumably at lower ambient densities and pressures. This is the second specific frequency problem.

Figure 2 illustrates this situation in M87, the cD galaxy at the center of the Virgo Cluster and the first system for which the effect was shown convincingly to exist (Harris 1986). The solid line in the top panel is the galaxy light profile, derived from the surface photometry of de Vaucouleurs & Nieto (1978), and the points trace the projected GCS number densities (in units of pc$^{-2}$ and scaled up for a direct comparison with the stellar densities) from the combined data of McLaughlin et al. (1993), McLaughlin (1995), and Harris (1986) (see McLaughlin 1999). It is clear that the radial gradient of $N_{cl}$ is significantly shallower than that of $I_V$, leading directly to the strongly increasing local specific frequency profile in the bottom panel.

The horizontal line in the bottom of Fig. 2 marks the globally averaged $S_N$ for M87 as a whole: $14.1 \pm 1.6$, three times higher than “normal” for giant ellipticals (Harris et al. 1998). That is, M87 also suffers from the first specific frequency problem. This is one strong hint that the two $S_N$ problems are really just different aspects of a single basic phenomenon. What this might be has become clear only with a homogeneous survey of the central galaxies (simply BCGs hereafter) in 21 Abell clusters by Blakeslee (1997), Blakeslee et al. (1997), and Blakeslee (1999).

2.2. X-RAY GAS AND A UNIVERSAL $\epsilon_{cl}$

Blakeslee has found (see also West et al. 1997) that the global specific frequencies of BCGs increase systematically with the soft X-ray luminosity (averaged over $\sim$500-kpc
scales) of the hot gas in their parent clusters. He then uses the details of this correlation to argue that the number of globulars in the cores of galaxy clusters scales in direct proportion to the total mass there, with one found for every $1–2 \times 10^{10} M_{\odot}$ of stars, gas, and dark matter. This implies that the first $S_N$ problem stems (in BCGs, at least) from a tendency for brighter galaxies to be “underluminous,” for the amount of gas and dark matter associated with them, rather than overabundant in globular clusters. Strictly from a star-formation point of view, however, only the baryons are of interest; thus, Harris et al. (1998) suggest, in essence, that if the dark matter were left out, the global mass ratio

$$\hat{\epsilon}_{cl} = \frac{M_{gcs}}{(M_{gas} + M_{stars})}$$

might itself be constant, not only among BCGs but in other galaxies as well. The first specific frequency problem would still arise more or less as Blakeslee suggested, with the (observed) larger gas fractions $M_{gas}/M_{stars}$ in brighter galaxies resulting also in a higher global $S_N \propto M_{gcs}/M_{stars}$. In addition, McLaughlin (1999) notes that if the local $\hat{\epsilon}_{cl}$—defined in the obvious way as a ratio of densities—were also constant as a function of radius in galaxies, then the fact that the X-ray gas in ellipticals tends to be hotter and more spatially extended than the stellar distribution could cause the second specific frequency problem in gas-rich galaxies that also have a high global $S_N$.

The underlying idea here is, of course, that the present-day sum $(M_{gas} + M_{stars})$ should be a better indicator of the “initial” gas mass in large galaxies, and that equation [4] may therefore be more accurate than $S_N$ as an estimate of the true cluster formation efficiency in equation [4]. Admittedly, $\hat{\epsilon}_{cl}$ is still only an observable approximation to $\epsilon_{cl}$; indeed, in a hierarchical universe it may be difficult to say precisely what is meant by $M_{gcs}^{init}$ in the first place. Particularly in a large galaxy which may have accreted gas (and stars and globulars) over extended lengths of time, the observable $\hat{\epsilon}_{cl}$ must be viewed as a mass-weighted average over a complex evolutionary history including a potentially large number of galaxy interactions and many discrete star-formation episodes. However, the current ratio of $M_{gcs}$ to $(M_{gas} + M_{stars})$ does have an essentially universal value in old GCSs—including those in regular gE’s and early-type dwarfs as well as BCGs—arguing that it may be a good reflection after all of the real $\epsilon_{cl}$.

Figures 3 shows this first for the local $\hat{\epsilon}_{cl}$ in M87 and in the bright ellipticals M49 (also in the Virgo Cluster) and NGC 1399 (BCG in the Fornax Cluster): The ratio of projected GCS mass densities ($\Sigma_{cl} \equiv \langle m \rangle N_{cl}$) to the sum of stellar and gas mass densities ($T_{V,gal} \Sigma V + \Sigma_{gas}$, with $T_{V,gal}$ measured separately for each system) is constant beyond an effective radius in the galaxies. The construction of the individual density profiles is discussed in detail by McLaughlin (1999), where a comparison of the deprojected quantities is also made, confirming the basic result and giving essentially the same numbers for $\hat{\epsilon}_{cl}$. Evidently, including the gas does alleviate the second specific frequency problem, in just the sense suggested above: $S_N$ increases with radius in M87 because of a locally varying gas-to-star mass ratio, rather than any change in the fundamental $\hat{\epsilon}_{cl}$; and in the comparatively gas-poor M49 and NGC 1399, there is no $S_N$ problem in the first place. Moreover, the first $S_N$ problem is similarly removed when the X-ray gas is taken into account: Although the total specific frequencies of M87, M49, and NGC 1399 are significantly different (at 14.1, 4.7, and 6–7), their local GCS mass ratios at large radii are consistent with a single value: $\hat{\epsilon}_{cl} = 0.0026 \pm 0.0005$ in the mean. Finally, since this is also

1 As Blakeslee (1999) discusses at length, his ratio of globulars per unit total mass and the ratio of equation [4] can both be constants in BCGs if the baryon fraction in the cores of galaxy clusters is also roughly universal (on the order of 10%). But as reasonable as it may seem, this possibility is in general unproven, and it is not obvious a priori that the two efficiencies are necessarily equivalent. Also, the constancy of Blakeslee’s ratio has been demonstrated neither globally for objects other than BCGs nor locally as a function of position inside any one galaxy. The behavior of $\hat{\epsilon}_{cl}$ in general can therefore not be anticipated from Blakeslee’s work.

2 The departure of the GCS densities from the stellar profiles at smaller radii in M87 and M49 is undoubtedly significant, but it is not clear whether this is due to a real decrease in the true cluster formation efficiency there, or to dynamical depletions of the initial GCS, or to substantial dissipation in the gas that formed the field stars (after the globulars were already in place) in the innermost regions of the galaxies.
independent of galactocentric radius, it is the same as the global cluster formation efficiency in each case. These specific examples clearly are consistent with the notion of a universal \( \epsilon_{cl} \). Figure 4, which essentially replots the \( S_N \) data in Fig. 1, confirms it in detail on a global scale. To understand the bold curves in this Figure, note that equation (4) and the definition \( N_{tot} = M_{gcs}/(2.4 \times 10^8 M_\odot) \) can be written as

\[
N_{tot} = 4.17 \times 10^6 \epsilon_{cl} \left( 1 + \frac{M_{gas}}{M_{stars}} \right) \left( \frac{M_{stars}}{10^{12} M_\odot} \right).
\]  

The heavy solid line in Fig. 4 is just this equation, with (i) \( \epsilon_{cl} = 0.0026 \) fixed; (ii) \( V \)-band galaxy luminosities converted to stellar masses according to the relation \( \Upsilon_{\text{V,gal}} = 6.3 M_\odot L_\odot^{-1} (L_{\text{V,gal}}/10^{11} L_\odot)^{0.9} \) (van der Marel 1991); and (iii) galaxy-wide gas-to-star mass ratios estimated from a combination of fundamental-plane scalings and the X-ray–optical luminosity correlation: \( M_{gas}/M_{stars} \approx 0.55 \times (L_{\text{V,gal}}/10^{11} L_\odot)^{1.5} \) (McLaughlin 1999). The steep upturn in \( N_{tot} \) at high \( L_{\text{V,gal}} \) (or the sharp increase in BCG specific frequency) is thus due to the fast-growing dominance of gas over stars, in the face of a constant GCS mass fraction. Towards lower luminosities, global gas masses become negligible and \( S_N \) decreases steadily because of the systematic decrease in \( \Upsilon_{\text{V,gal}} \) for fundamental-plane ellipticals. As a result, at the luminosity of the Milky Way spheroid (disk excluded), an \( \epsilon_{cl} \) of 0.0026 also accounts for the number of halo (metal-poor) Galactic globulons (open square in Fig. 4; see McLaughlin 1999).

For the early-type dwarf galaxies with \( L_{\text{V,gal}} \leq 2 \times 10^9 L_\odot \), the gas-to-star mass ratio in equation (4) has a different meaning: The energy of supernova explosions in a single burst of star formation in one of these small galaxies may have sufficed to expel all remaining gas from its dark-matter well, and while any such gas would, of course, no longer be directly observable, a proper estimate of \( \epsilon_{cl} \) must still account for it. The bold, dashed line in Fig. 4 is one attempt to do this. It represents equation (5) given (i) the relation \( (1 + M_{gcs}/M_{stars}) \sim (L_{\text{V,gal}}/2 \times 10^9 L_\odot)^{-0.4} \), from the theory of Dekel & Silk (1986), (ii) a constant \( \Upsilon_{\text{V,gal}} = 2 M_\odot L_\odot^{-1} \) for the stellar populations; and (iii) once again, a fixed \( \epsilon_{cl} = 0.0026 \).

Although scatter remains (at the level of factors of \( \sim 2 \)) in the observed \( N_{tot} \) at any given \( L_{\text{V,gal}} \) in Fig. 4, it is important that this is more or less random; the mean trends in GCS population as a function of luminosity—the essence of the first specific frequency problem—can be simply explained if the efficiency of globular cluster formation were constant to first order. Indeed, deviations in Fig. 4 may reflect the scatter of individual galaxies about either the fundamental plane or the \( L_X-L_B \) correlation used to derive the bold lines there, rather than any significant variations in \( \epsilon_{cl} \). This requires further study on a case-by-case basis, as does the situation in spirals other than the Milky Way. Similarly, there is some indication (e.g., Mac Low & Ferrara 1999) that the simple treatment of galactic winds (i.e., the model of Dekel & Silk 1986) used to correct for the gas lost from dwarf galaxies may be inadequate. At this point, however, it is more than plausible to assert that globular clusters formed in dEs and dSph’s as in larger galaxies, always in the same proportion to the total mass of gas that was initially on hand.

One important consequence of this is that the efficiency of unclustered star formation in protogalaxies could not have been universal. In both the faintest dEs and the brightest BCGs, globular clusters apparently formed in precisely the numbers expected of them, while anomalously low fractions of the initial gas mass were converted into field stars. In the case of the dwarfs specifically, if even just the idea of the feedback correction above is basically correct then all the globulars had to have formed by the time a galactic wind cleared the remaining gas; but this must have happened before normal numbers of field stars appeared. Thus, the gas which formed bound star clusters had to have collapsed more rapidly than that which produced unbound groups and associations. This implies that it was only those pieces of gas which locally exceeded some critical density that were able to attain the cumulative star formation efficiency of \( \eta \approx 20\%–50\% \) required to form a bound stellar cluster. In addition to this, the uniformity of \( \epsilon_{cl} \) argues—applying as it does over large ranges of radius inside M87, M49, and NGC 1399, and from dwarfs in the field to BCGs in the cores of Abell clusters—that the probability of realizing such a high SFE depended very weakly, if at all, on local or global protogalactic environment. Quantitative theories of cluster formation should therefore seek to identify a threshold in relative density,
\( \delta \rho / \rho \), that is always exceeded by \( \gtrsim 0.26\% \) of the mass fluctuations in any large body of star-forming gas.

The “relative” aspect of such a criterion is crucial; the GCS data militate strongly against any model relying on parameters that are too sensitive to environment. One such example is the scenario of Elmegreen & Efremov (1997) in which the pressure exerted by a diffuse medium surrounding a dense clump of gas must exceed a fixed, absolute value in order to produce a high local \( \eta \) and a bound stellar cluster. However, since pressures vary by orders of magnitude in going from dE’s to BCGs, or from large to small radii in any one galaxy, this idea seems to imply systematic variations in \( \epsilon_{\text{cl}} \) that are not observed.

BCGs present a complex problem in larger-scale galaxy formation, but it is worth noting that a feedback argument like that applied to dwarfs may also be relevant to central cluster galaxies like M87 (cf. Harris et al. 1998 [McLaughlin 1999]). That is, globulars likely also appeared quickly, and in normal numbers, in the densest of star-forming clumps (perhaps embedded in dwarf-sized fragments) in these deep potential wells. The gas more slowly forming field stars could have been virialized thereafter, or moved outwards in slow, partial galactic winds. The unused gas in this case would have to remain hot to the present day, and more or less in the vicinity of the parent galaxy, in order to appear as the X-ray emitting gas that makes \( \epsilon_{\text{cl}} \) so constant in Fig. [3], but this requirement is certainly consistent with the BCGs being at the centers of clusters. In addition, the feedback in this scenario would have more effectively truncated the star formation in the lower-density environs at larger galactocentric radii in these very large systems, thus giving rise to the second specific frequency problem as well. There are other possibilities for BCGs, however. It is conceivable, for instance, that their “excess” gas and globulars were both produced elsewhere in galaxy clusters (in failed dwarfs?) and fell together onto the central galaxies over a long period of time. These questions need to be examined in much more detail.

Finally, McLaughlin (1999) argues that the current efficiency of open cluster formation in the Galactic disk is also \( \sim 0.2–0.4\% \) by mass. This figure is much more uncertain than it is in GCSs, and it is essentially an instantaneous variant of the time-averaged quantity measured for globular clusters. Nevertheless, it clearly suggests that whatever quantitative criterion is ultimately required to explain \( \epsilon_{\text{cl}} = 0.26\% \) in GCSs may very well prove to be of much wider applicability. (One exception may be the formation of massive clusters in mergers and starbursts, where it has been suggested that \( \epsilon_{\text{cl}} \sim 1–10\% \) [e.g., Zepf et al. 1999; Schweizer 1999]. However, this conclusion is very uncertain and requires more careful investigation.)

3. Globular Cluster Binding Energies

The focus to this point has been on the frequency with which \( \sim 10^5–10^6 M_\odot \) clumps of gas were able to form stars with a cumulative efficiency \( \eta \) high enough to produce a bound globular cluster. The impressive regularity of this occurrence is clearly important, as has just been discussed, and its rarity is significant as well: the small value of \( \epsilon_{\text{cl}} = 0.26\% \) implies that the local SFE in an average bit of protogalactic gas was much lower than \( \eta_{\text{crit}} \sim 0.2–0.5 \) (a fact which is also true of molecular gas in the Galaxy today). However, these results say nothing of how an extreme \( \eta \) comes about in any individual gas clump. This is another open problem in star formation generally. Its solution requires both an understanding of local star formation laws \( (d \rho_\text{g} / dt) \) as a function of \( \rho_\text{gas} \) and a self-consistent treatment of feedback on small (\( \sim 10–100 \) pc) scales.

The whole issue is essentially one of energetics in a compact, gravitationally bound association of gas and embedded young stars: When does the combined energy injected by all the massive stars equal the binding energy of whatever gas remains? This point of equality and the corresponding \( \eta \) can in principle be identified for any given star formation law and a set of initial conditions in the original gas. The difficulty lies, of course, in deciphering what these are; but once this is done, progress will also have been made in understanding the probability of obtaining \( \eta > \eta_{\text{crit}} \), i.e., the overall efficiency of cluster formation in \( \eta_{\text{cl}} \). One way to begin addressing this complex set of problems empirically is to compare the final binding energies of stellar clusters with the initial energies of their gaseous progenitors. This is a straightforward exercise for the ensemble of globular clusters in the Milky Way.

Saito (1979a) evaluated the binding energies \( E_b \) for about 10 bright Galactic globulars, along with a number of dwarf and giant ellipticals. He claimed that \( E_b \propto M^{1.5} \) for dEs and globulars alike, while the dwarfs fell systematically below this relation (a fact which he subsequently attributed to the effects of large-scale feedback such as discussed above [Saito 1979b, but cf. Bender et al. 1992]). Twenty years later, the data required for a calculation of binding energy are available for many more than ten globular clusters, and they are of higher quality than those available to Saito.

McLaughlin (2000) computes the binding energies of 109 “regular” Galactic globulars and 30 objects with post–core-collapse (PCC) morphologies. The main assumption is that single-mass, isotropic King (1966) models provide a complete description of the clusters’ internal structures. Within this framework, the fundamental definition \( E_b \equiv -1/(2) \int_0^{r_t} 4\pi r^2 \rho \, dr \) (with \( r_t \) the tidal radius of the cluster) may be written as

\[
E_b = 1.663 \times 10^{44} \text{ erg} \left( \frac{r_0}{\text{pc}} \right)^5 \left( \frac{\Upsilon_{V,0} j_0}{M_\odot \text{ pc}^{-3}} \right)^2 \mathcal{E}(c), \tag{6}
\]

where \( r_0 \) is the model scale radius (King 1966), \( \Upsilon_{V,0} \) the core mass-to-light ratio; \( j_0 \) is the central V-band luminosity density; and \( \mathcal{E}(c) \) is a well defined, nonlinear function of \( c \equiv \log (r_t/r_0) \), obtained by numerically integrating King models (McLaughlin 2000).
Values of $r_0$, $j_0$, and $c$ are given for all Milky Way clusters in the catalogue of Harris (1996). However, a determination of $\Upsilon_{V,0}$ requires measurements of velocity dispersions $\sigma_0$, and these are available for only a third of the sample, in the compilation of Pryor & Meylan (1993). For these objects, application of the King-model relation $\Upsilon_{V,0} = 9\sigma_0^2/(4\pi G r_0^2 j_0)$ gives the results shown in the top panel of Fig. 5. The regular globulars there (the 39 filled circles) share a single, constant core mass-to-light ratio: $\langle \log \Upsilon_{V,0} \rangle = 0.16 \pm 0.03$ in the mean, and the r.m.s. scatter about this is less than the 1-σ observational errorbar shown for $\log \Upsilon_{V,0}$. (The results for 17 PCC clusters, plotted as open squares, are almost certainly spurious [see McLaughlin 2000]. They are shown for completeness but not included in any quantitative analyses here.) This is consistent with separate work by Mandushev et al. (1991) and Pryor & Meylan (1993).

It can safely be assumed that this same $\Upsilon_{V,0}$ applies to all other (non-PCC) Galactic globulars, so that $E_b$ can be computed from equation (6) given only $r_0$, $j_0$, and $c$, i.e., on the basis of cluster photometry or star-count data alone. If this is done for the full Harris (1996) catalogue, a very tight correlation between binding energy, total cluster luminosity, and Galactocentric position is found:

$$E_b = 7.2 \times 10^{39} \text{ erg} \ (L/ L_\odot)^{2.05} (r_{gc}/8 \text{ kpc})^{-0.4},$$

(7)

with uncertainties of about $\pm 0.1$ in the fitted powers on $L$ and $r_{gc}$. This relation is drawn as the line in the middle panel of Fig. 5. The r.m.s. scatter of the regular-cluster data (filled and open circles) about it is no larger than the typical 1-σ observational uncertainty on $\log E_b$.

So far as current data can tell, the constancy of $\Upsilon_{V,0}$ and the scaling of $E_b$ with $L$ and $r_{gc}$ are essentially perfect. Now, in the context of King (1966) models, any globular cluster is fully defined by the specification of just four (nominally) independent physical quantities. Given the results just presented, it is natural to choose these to be $\log \Upsilon_{V,0}$, $\log E_b$, the total $\log L$, and the concentration parameter $c = \log (r_t/r_0)$. (Additional factors such as Galactocentric position or cluster metallicity are quite separate from the model characterization of a cluster, and they are to be viewed as external parameters.) But the tight empirical constraints on $\Upsilon_{V,0}$ and $E_b$ mean that they are not actually “free” in any real sense; in practice, Galactic globulars are only a two-parameter family, with all internal properties set by $\log L$ and $c$. Equivalently, the clusters are confined to a fundamental plane in the larger, four-dimensional space of King models available to them in principle. The top plots in Fig. 5 are then just two edge-on views of this plane. Its properties are discussed in detail by McLaughlin (2000) where this physical approach to it is also compared to the more statistical tack taken by Djorgovski (1995) who first claimed its existence, and to the different interpretation suggested by Bellazzini (1998).

The bottom panel shows the third plot possible in the physical cluster “basis” chosen here: concentration vs. total.

Figure 5. The fundamental plane of Galactic globular clusters (after McLaughlin 2000). Top panels are two edge-on views; bottom is nearly the face-on view. All correlations between any other combinations of cluster observables follow directly from these three relations between $\Upsilon_{V,0}$, $E_b$, $c$, $L$, and $r_{gc}$.
the observations errorbar on $c$. Neither the slope nor the normalization of this rough correlation changes with Galactocentric position, i.e., the distribution of globulars on the fundamental plane is independent of $r_{gc}$.

The mean core mass-to-light ratio is also independent of Galactocentric radius, and none of the distributions in Fig. 3 depend on cluster metallicity. Moreover, since any other property of a King-model cluster is known once values for $\Upsilon_{V,0}$, $E_b$, $c$, and $L$ are given, it follows that all interdependences between any globular cluster observables (and there are many; see, e.g., Djorgovski & Meylan 1994) are perforce equivalent to a combination of (i) a constant $\Upsilon_{V,0} = 1.45\,M_\odot/L_\odot^{-1}$; (ii) equation (7) for $E_b$ as a function of $L$ and $r_{gc}$; (iii) the rough increase of $c$ with $L$; and (iv) generic King-model definitions. McLaughlin (2000) derives a complete set of structural and dynamical correlations to confirm this basic point: only the quantitative details of Fig. 3 and their insensitivity to metallicity—need be explained in any theory of globular cluster formation and evolution in our Galaxy.

It is then important that the $E_b(L, r_{gc})$ and $c(L)$ correlations are stronger among clusters outside the Solar circle (filled circles in the plots) than among those within it (open circles). Given the relative weakness of dynamical evolution at such large $r_{gc}$, this is one indication that these fundamental properties of the Galactic GCS were set largely by the cluster formation process (see also Murray & Lin 1992; Bellazzini et al. 1996; Vesperini 1997).

Figure 6 finally compares the globular cluster energies to estimates for the initial values in their progenitors. This is done for the specific model of Harris & Pudritz (1994) in which protoglobular clusters are embedded in larger protogalactic fragments and have properties analogous to those of the dense clumps inside present-day molecular clouds (see [5]). In particular, the column densities of the protoclusters are postulated to be independent of mass but decreasing with Galactocentric radius: $M/\pi R^2 \approx 10^3 M_\odot$ pc$^{-2}$ ($r_{gc}/8\text{kpc})^{-1}$, which follows from the clumps being in hydrostatic equilibrium and from their parent clouds being themselves surrounded in a diffuse medium virialized in a “background” isothermal potential well with a circular velocity of 220 km s$^{-1}$. This relation then implies $E_b \equiv GM^2/R = 4.8 \times 10^{42}$ erg ($M/M_\odot)^{1.5}$ ($r_{gc}/8\text{kpc})^{-0.5}$, which is drawn as the broken line in Fig. 6. By construction, this is precisely the mass-energy relation obeyed today by the massive clumps in molecular clouds in the Solar neighborhood. Intriguingly, it is also the $M - E_b$ scaling originally claimed by Saito (1970a) for giant elliptical galaxies and (bright) globular clusters.

The dependence of $E_b$ on $r_{gc}$ in such protoclusters is nearly the same as that actually found for the globulark today. (It similarly accounts for the observed increase of cluster radii with $r_{gc}$ Harris & Pudritz 1994, cf. Murray & Lin 1992—a trend which is, in fact, equivalent to the behavior of $E_b$ in eq. 7.) In Fig. 6, the two are taken for convenience to be identical, so that the comparison between model and observed binding energies there is valid at any given Galactocentric position.

The difference in the slopes of the two $E_b(M)$ relations is significant: The ratio of the initial energy of a gaseous clump to the final $E_b$ of a stellar cluster is unavoidably a function of its cumulative star formation efficiency $\eta$; but Fig. 6 shows that this ratio of energies changes systematically as a function of mass, and thus that $\eta$ varied as well. Moreover, the fact that the difference between initial and final $E_b$ is largest at the lowest masses implies that $\eta$ had to have been lower in lower-mass protoglobulars. The details of this behavior must rely on the density and velocity structure of the initial gas; the timescale over which feedback expels unused gas; re-expansion of the stars after such gas loss; and other such specifics which are model-dependent to some extent. The inference on the qualitative behavior of $\eta$ as a function of protocluster gas mass is, however, robust.

A more quantitative discussion of Fig. 6— including its implications for the mass function of GCs, which will differ from the mass functions of gaseous protoclusters if $\eta$ varied systematically from one to the other—has to be deferred (McLaughlin, in progress). But this evidence for a variable star formation efficiency in protoclusters is itself a new target for theoretical attack, most likely through a
general calculation of star formation and feedback such as that described at the beginning of this Section. As was also mentioned there, if such models can be made to agree with the data in Fig. 3 they will likely also shed new light on the empirical efficiency of cluster formation, and perhaps on other generic properties of GCSs as well—and, thence, on larger-scale processes in galaxy formation.

Acknowledgements
This work was supported by NASA through grant number HF-1097.01-97A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA under contract NAS5-26555.

References
Aguilar, L., Hut, P., & Ostriker, J. P. 1988, ApJ, 335, 720
Ashman, K. M., & Zepf, S. E. 1992, ApJ, 384, 50
Ashman, K. M., & Zepf, S. E. 1998, Globular Cluster Systems (New York: Cambridge Univ. Press)
Bellazzini, M. 1998, New Astronomy, 3, 219
Bellazzini, M., Vesperini, E., Ferraro, F. R., & Fusi Pecci, F. 1996, MNRAS, 279, 337
Bender, R., Burstein, D., & Faber, S. 1992, ApJ, 399, 462
Blakeslee, J. P. 1997, ApJ, 481, L59
Blakeslee, J. P. 1999, AJ, 118, 1506
Blakeslee, J. P., Tonry, J. L., & Metzger, M. R. 1997, AJ, 114, 482
Caputo, F., & Castellani, V. 1984, MNRAS, 207, 185
Capuzzo-Dolcetta, R. 1993, ApJ, 415, 616
Côté, P., Marzke, R. O., & West, M. J. 1998, ApJ, 501, 554
Côté, P., Marzke, R. O., West, M. J., & Minniti, D. 2000, ApJ, in press [astro-ph/9911527]
de Vaucouleurs, G., & Nieto, J.-L. 1978, ApJ, 220, 449
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Djorgovski, S., & Meylan, G. 1994, AJ, 108, 1292
Durrell, P. R., Harris, W. E., Geisler, D., & Pudritz, R. E. 1996, AJ, 112, 972
Elmegreen, B. G. 2000, in Toward a New Millennium in Galaxy Morphology, ed. D. L. Block, I. Puerari, A. Stockton, and D. Ferreira (Dordrecht: Kluwer), in press [astro-ph/9911157]
Elmegreen, B. G., & Clemens, C. 1985, ApJ, 294, 523
Elmegreen, B. G., & Efremov, Y. N. 1997, ApJ, 480, 235
Fall, S. M., & Rees, M. J. 1977, MNRAS, 181, 37P
Fall, S. M., & Rees, M. J. 1985, ApJ, 298, 18
Forbes, D. A., Brodie, J. P., & Grillmair, C. J. 1997, AJ, 113, 1652
Gnedin, O. Y., & Ostriker, J. P. 1997, ApJ, 474, 223
Goodwin, S. P. 1997, MNRAS, 284, 785
Harris, W. E. 1986, AJ, 91, 822
Harris, W. E. 1991, ARA&A, 29, 543
Harris, W. E. 1996, AJ, 112, 1487
Harris, W. E. 2000, Globular Cluster Systems: Lectures for the 1998 Saas-Fee Advanced Course on Star Clusters (Swiss Society for Astrophysics and Astronomy), in press (preprint at http://physun.mcmaster.ca/~harris/WEHarris.html)
Harris, W. E., & van den Bergh, S. 1981, AJ, 86, 1627
Harris, W. E., Harris, G. L. H., & McLaughlin, D. E. 1998, AJ, 115, 1801
Hills, J. G. 1980, ApJ, 225, 986
King, I. R. 1966, AJ, 71, 64
Kissler-Patig, M. 1997, A&A, 319, 83
Kissler-Patig, M., & Gebhardt, K. 1999, AJ, 118, 1526
Kissler-Patig, M., Forbes, D. A., & Minniti, D. 1998, MNRAS, 298, 1123
Lada, E. A., DePoy, D. L., Evans, N. J., & Gatley, I. 1991, ApJ, 371, 171
Lada, E. A. 1992, ApJ, 393, L25
Larson, R. B. 1988, in IAU Symposium No. 126, Globular Cluster Systems in Galaxies, ed. J. E. Grindlay and A. G. D. Philip (Dordrecht: Kluwer), 311
Larson, R. B. 1993, in ASP Conf. Ser. 48, The Globular Cluster–Galaxy Connection, ed. G. H. Smith and J. P. Brodie (San Francisco: ASP), 675
Mac Low, M.-M., & Ferrara, A. 1999, ApJ, 513, 142
Mandushev, G., Spassova, N., & Staneva, A. 1991, A&A, 252, 94
Mathieu, R. D. 1983, ApJ, 267, L97
McLaughlin, D. E. 1995, AJ, 109, 2034
McLaughlin, D. E. 1999, AJ, 117, 2398
McLaughlin, D. E. 2000, ApJ, in press
McLaughlin, D. E., & Pudritz, R. E. 1996, ApJ, 457, 578
McLaughlin, D. E., Harris, W. E., & Hanes, D. A. 1993, ApJ, 409, L45
Meurer, G. R., Heckman, T. M., Leitherer, C., Kinney, A., Robert, C., & Garnett, D. R. 1995, AJ, 110, 2665
Meylan, G., & Heggie, D. C. 1997, A&A Rev., 8, 1
Miller, B. W., Lotz, J. M., Ferguson, H. C., Stiavelli, M., & Whitmore, B. C. 1998, ApJ, 508, L133
Murali, C., & Weinberg, M. D. 1997a, MNRAS, 288, 749
Murali, C., & Weinberg, M. D. 1997b, MNRAS, 288, 767
Murali, C., & Weinberg, M. D. 1997c, MNRAS, 291, 717
Murray, S. D., & Lin, D. N. C. 1992, ApJ, 400, 265
Murray, S. D., & Lin, D. N. C. 1993, in The Globular Cluster–Galaxy Connection, ed. G. H. Smith and J. P. Brodie (San Francisco: ASP), 738
Ostriker, J. P., Binney, J., & Saha, P. 1989, MNRAS, 241, 849
Pate1, K., & Pudritz, R. E. 1994, ApJ, 424, 688
Peebles, P. J. E., & DeEicke, R. H. 1968, ApJ, 150, 891
Pryor, C., & Meylan, G. 1993, in ASP Conf. Ser. 50, Structure and Dynamics of Globular Clusters, ed. S. G. Djorgovski and G. Meylan (San Francisco: ASP), 357
Saito, M. 1979a, PASJ, 31, 181
Saito, M. 1979b, PASJ, 31, 193
Schneider, F. 1987, in Nearly Normal Galaxies, ed. S. M. Faber (New York: Springer), 18
Schneider, F. 1999, in Spectrophotometric Dating of Stars and Galaxies, ed. I. Hubeny, S. Heap, and R. Cornett (San Francisco: ASP), in press [astro-ph/9906488]
Seelur, L., & Zinn, R. 1978, ApJ, 225, 357
Spitzer, L. 1987, Dynamical Evolution of Globular Clusters (Princeton: Princeton Univ. Press)
van den Bergh, S. 1994, ApJ, 435, 203
van der Marel, R. P. 1991, MNRAS, 253, 710
Verschueren, W. 1990, A&A, 234, 156
Vesperini, E. 1997, MNRAS, 287, 915
West, M. J., Côté, P., Jones, C., Forman, W., & Marzke, R. O. 1995, ApJ, 453, L77
Whitmore, B. C., & Schweizer, F. 1995, AJ, 109, 960
Zepf, S. E., & Ashman, K. M. 1993, MNRAS, 264, 611
Zepf, S. E., Ashman, K. M., English, J., Freeman, K. C., & Sharples, R. M. 1999, AJ, 118, 752