1 Globular Cluster Systems

William E. Harris

Department of Physics & Astronomy, McMaster University
Hamilton ON L8S 4M1, Canada

Abstract. 1. The basic features of the globular cluster system of the Milky Way are summarized: the total population, subdivision of the clusters into the classic metal-poor and metal-rich components, and first ideas on formation models. The distance to the Galactic center is derived from the spatial distribution of the inner bulge clusters, giving $R_0 = (8.1 \pm 1.0)$ kpc.

2. The calibration of the fundamental distance scale for globular clusters is reviewed. Different ways to estimate the zero point and metallicity dependence of the RR Lyrae stars include statistical parallax and Baade-Wesselink measurements of field RR Lyraes, astrometric parallaxes, white dwarf cluster sequences, and field subdwarfs and main sequence fitting. The results are compared with other distance measurements to the Large Magellanic Cloud and M31.

3. Radial velocities of the Milky Way clusters are used to derive the kinematics of various subsamples of the clusters (the mean rotation speed about the Galactic center, and the line-of-sight velocity dispersion). The inner metal-rich clusters behave kinematically and spatially like a flattened rotating bulge population, while the outer metal-rich clusters resemble a thick-disk population more closely. The metal-poor clusters have a significant prograde rotation (80 to 100 km s^{-1}) in the inner halo and bulge, declining smoothly to near-zero for $R \gtrsim R_0$. No identifiable subgroups are found with significant retrograde motion.

4. The radial velocities of the globular clusters are used along with the spherically symmetric collisionless Boltzmann equation to derive the mass profile of the Milky Way halo. The total mass of the Galaxy is near $\simeq 8 \times 10^{11} M_\odot$ for $r \lesssim 100$ kpc. Extensions to still larger radii with the same formalism are extremely uncertain because of the small numbers of outermost satellites, and the possible correlations of their motions in orbital families.

5. The luminosity functions (GCLF) of the Milky Way and M31 globular clusters are defined and analyzed. We search for possible trends with cluster metallicity or radius, and investigate different analytic fitting functions such as the Gaussian and power-law forms.

6. The global properties of GCSs in other galaxies are reviewed. Measureable distributions include the total cluster population (quantified as the specific frequency S_N), the metallicity distribution function (MDF), the luminosity and space distributions, and the radial velocity distribution.

7. The GCLF is evaluated as a standard candle for distance determination. For giant E galaxies, the GCLF turnover has a mean luminosity of $M_V = -7.33$ on a distance scale where Virgo has a distance modulus of 31.0 and Fornax is at 31.3, with galaxy-to-galaxy scatter $\sigma(M_V) = 0.15$ mag. Applying this calibration to more remote galaxies yields a Hubble constant $H_0 = (74 \pm 9)$ km s^{-1} Mpc^{-1}.

8. The observational constraints on globular cluster formation models are summarized. The appropriate host environments for the formation of $\sim 10^5 - 10^6 M_\odot$ clusters are suggested to be kiloparsec-sized gas clouds (Searle/Zinn fragments or supergiant molecular clouds) of $10^{8-9} M_\odot$. A model for the growth of protocluster clouds by collisional agglomeration is presented, and matched with observed mass distribution functions. The issue of globular cluster formation efficiency in different galaxies is discussed (the "specific frequency problem").

9. Other influences on galaxy formation are discussed, including mergers, accretions, and starbursts. Mergers of disk galaxies almost certainly produce elliptical galaxies of *low* S_N, while the high$-S_N$ ellipticals are more likely to have been produced through *in situ* formation. Starburst dwarfs and large active galaxies in which current globular cluster formation is taking place are compared with the key elements of the formation model.

10. (Appendix) Some basic principles of photometric methods are gathered together and summarized: the fundamental signal-to-noise formula, objective star finding, aperture photometry, PSF fitting, artificial-star testing, detection completeness, and photometric uncertainty. Lastly, we raise the essential issues in photometry of nonstellar objects, including image moment analysis, total magnitudes, and object classification techniques.

INTRODUCTION

When the storm rages and the state is threatened by shipwreck, we can do nothing more noble than to lower the anchor of our peaceful studies into the ground of eternity.

Johannes Kepler

Why do we do astronomy? It is a difficult, frustrating, and often perverse business, and one which is sometimes costly for society to support. Moreover, if we are genuinely serious about wanting to probe Nature, we might well employ ourselves better in other disciplines like physics, chemistry, or biology, where at least we can exert experimental controls over the things we are studying, and where progress is usually less ambiguous.

Kepler seems to have understood why. It is often said that we pursue astronomy because of our inborn curiosity and the need to understand our place and origins. True enough, but there is something more. Exploring the universe is a unique adventure of profound beauty and exhilaration, lifting us far beyond our normal self-centered concerns to a degree that no other field can do quite as powerfully. Every human generation has found that the world beyond the Earth is a vast and astonishing place.

In these chapters, we will be taking an all-too-brief tour through just one small area of modern astronomy – one which has roots extending back more than a century, but which has re-invented itself again and again with the advance of both astrophysics and observational technology. It is also one which draws intricate and sometimes surprising connections among stellar populations, star formation, the earliest history of the galaxies, the distance scale, and cosmology.

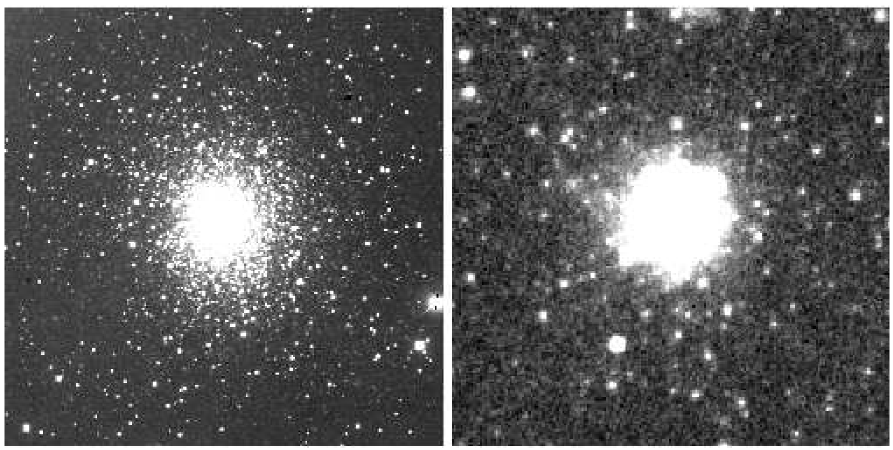

Fig. 1.1. *Left panel:* Palomar 2, a globular cluster about 25 kiloparsecs from the Sun in the outer halo of the Milky Way. The picture shown is an $I-$band image taken with the Canada-France-Hawaii Telescope (see Harris et al. 1997c). The field size is 4.6 arcmin (or 33 parsecs) across and the image resolution ("seeing") is 0.5 arcsec. *Right panel:* A globular cluster in the halo of the giant elliptical galaxy NGC 5128, 3900 kiloparsecs from the Milky Way. The picture shown is an $I-$band image taken with the Hubble Space Telescope (see G. Harris et al. 1998); the field size is 0.3 arcmin (or 340 parsecs) across, and the resolution is 0.1 arcsec. This is the most distant globular cluster for which a color-magnitude diagram has been obtained

The sections to follow are organized in the same way as the lectures given at the 1998 Saas-Fee Advanced Course held in Les Diablerets. Each one represents a well defined theme which could in principle stand on its own, but all of them link together to build up an overview of what we currently know about globular cluster systems in galaxies. It will (I hope) be true, as in any active field, that much of the material will already be superseded by newer insights by the time it is in print. Wherever possible, I have tried to preserve in the text the conversational style of the lectures, in which lively interchanges among the speakers and audience were possible. The literature survey for this paper carries up to the early part of 1999.

A *globular cluster system* (GCS) is the collection of all globular star clusters within one galaxy, viewed as a subpopulation of that galaxy's stars. The essential questions addressed by each section of this review are, in sequence:

- What are the size and structure of the Milky Way globular cluster system, and what are its definable subpopulations?
- What should we use as the fundamental Population II distance scale?
- What are the kinematical characteristics of the Milky Way GCS? Do its subpopulations show traces of different formation epochs?
- How can the velocity distribution of the clusters be used to derive a mass profile for the Milky Way halo?
- What is the luminosity ($\equiv$ mass) distribution function for the Milky Way GCS? Are there detectable trends with subpopulation or galactocentric distance?
- What are the overall characteristics of GCSs in other galaxies – total numbers, metallicity distributions, correlations with parent galaxy type?
- How can the luminosity distribution function (GCLF) be used as a "standard candle" for estimation of the Hubble constant?
- Do we have a basic understanding of how globular clusters formed within protogalaxies in the early universe?
- How do we see globular cluster populations changing today, due to such phenomena as mergers, tidal encounters, and starbursts?

The study of globular cluster systems is a genuine hybrid subject mixing elements of star clusters, stellar populations, and the structure and history of all types of galaxies. Over the past two decades, it has grown rapidly along with the spectacular advances in imaging technology. The first review article in the subject (Harris & Racine 1979) spent its time almost entirely on the globular clusters in Local Group galaxies and only briefly discussed the little we knew about a few Virgo ellipticals. Other reviews (Harris 1988a,b, 1991, 1993, 1995, 1996b, 1998, 1999) demonstrate the growth of the subject into one which can put a remarkable variety of constraints on issues in galaxy formation and evolution. Students of this subject will also want to read the recent book *Globular Cluster Systems* by Ashman & Zepf (1998), which gives another comprehensive overview in a different style and with different emphases on certain topics.

I have kept abbreviations and acronyms in the text to a minimum. Here is a list of the ones used frequently:

CMD: color-magnitude diagram
GCS: globular cluster system; the collection of all globular clusters in a given galaxy
GCLF: globular cluster luminosity function, conventionally defined as the number of globular clusters per unit *magnitude* interval $\phi(M_V)$

LDF: luminosity distribution function, conventionally defined as the number of globular clusters per unit *luminosity*, dN/dL. The LDF and GCLF are related through $\phi \sim L(dN/dL)$
MDF: metallicity distribution function, usually defined as the number of clusters (or stars) per [Fe/H] interval
MPC: "metal-poor component"; the low-metallicity part of the MDF
MRC: "metal-rich component"; the high-metallicity part of the MDF
ZAMS: zero-age main sequence; the locus of unevolved core hydrogen burning stars in the CMD
ZAHB: zero-age horizontal branch: the locus of core helium burning stars in the CMD, at the beginning of equilibrium helium burning

1.1 THE MILKY WAY SYSTEM: A GLOBAL PERSPECTIVE

It is a capital mistake to theorize before one has data.
Sherlock Holmes

We will see in the later sections that our ideas about the general characteristics of globular cluster systems are going to be severely limited, and even rather badly biased, if we stay only within the Milky Way. But the GCS of our own Galaxy is quite correctly the starting point in our journey. It is not the largest such system; it is not the most metal-poor or metal-rich; it is probably not the oldest; and it is certainly far from unique. It is simply the one we know best, and it has historically colored all our ideas and mental images of what we mean by "globular clusters", and (even more importantly) the way that galaxies probably formed.

1.1.1 A First Look at the Spatial Distribution

Currently, we know of 147 objects within the Milky Way that are called globular clusters (Harris 1996a). They are found everywhere from deep within the Galactic bulge out to twice the distance of the Magellanic Clouds. Fig. 1.2 shows the spatial distribution of all known clusters within ~ 20 kiloparsecs of the Galactic center, and (in an expanded scale) the outermost known clusters. To plot up these graphs, I have already assumed a "distance scale"; that is, a specific prescription for converting apparent magnitudes of globular cluster stars into true luminosities. As discussed in Section 2, this prescription is

$$M_V(HB) = 0.15\,[\mathrm{Fe/H}] \,+\, 0.80$$

where [Fe/H] represents the cluster heavy-element abundance (metallicity) and $M_V(HB)$ is the absolute V magnitude of the horizontal branch in the color-magnitude diagram (abbreviated CMD; see the Appendix for a sample cluster in which the principal CMD sequences are defined). For metal-poor clusters in which RR Lyrae stars are present, by convention $M_V(HB)$ is identical to the mean luminosity of these RR Lyraes. For metal-richer clusters in which there are only red HB stars and no RR Lyraes, $M_V(HB)$ is equal to the mean luminosity of the RHB. More will be said about the calibration of this scale in Section 2; for now, we will simply use it to gain a broad picture of the entire system.

Throughout this section, the numbers (X, Y, Z) denote the usual distance coordinates of any cluster relative to the Sun: X points from the Sun in toward the Galactic center, Y points in the direction of Galactic rotation, and Z points perpendicular to the Galactic plane northward. The coordinates (X, Y, Z) are defined as:

$$X = R \cos b \cos \ell\,, \quad Y = R \cos b \sin \ell\,, \quad Z = R \sin b\,, \tag{1.1}$$

where (ℓ, b) are the Galactic longitude and latitude and R is the distance of the cluster from the Sun. In this coordinate system the Sun is at $(0, 0, 0)$ kpc and the Galactic center at $(8, 0, 0)$ kpc (see below).

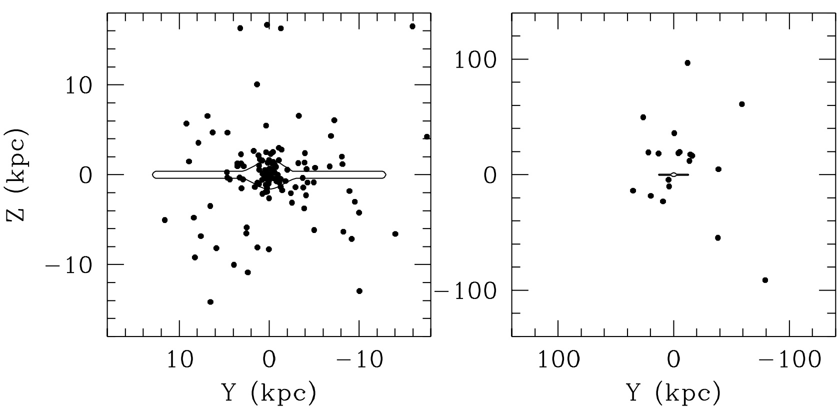

Fig. 1.2. *Left panel:* Spatial distribution of the inner globular cluster system of the Milky Way, projected onto the YZ plane. Here the Sun and Galactic center are at $(0, 0)$ and we are looking in along the X−axis toward the center. *Right panel:* Spatial distribution in the YZ plane of the outer clusters

In *very* rough terms, the GCS displays spherical symmetry – at least, as closely as any part of the Galaxy does. Just as Harlow Shapley did in the early part of this century, we still use it today to outline the size and shape of the Galactic *halo* (even though the halo field stars outnumber those in

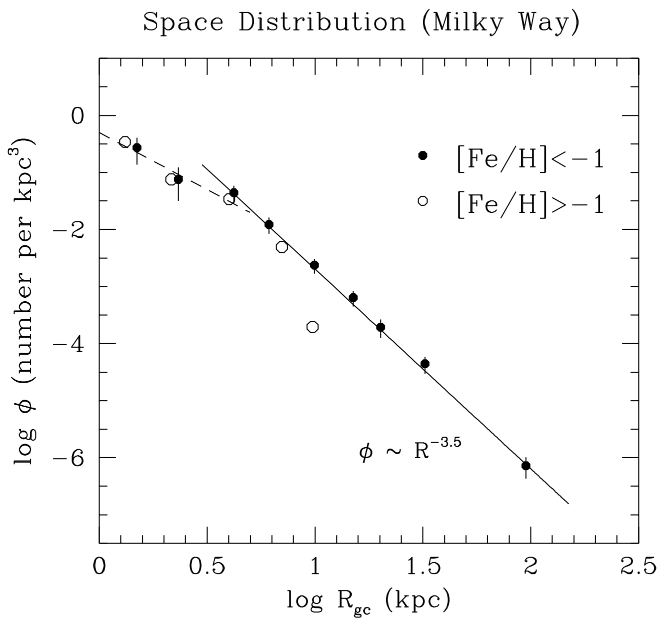

Fig. 1.3. Spatial distribution ϕ (number of clusters per unit volume) as a function of Galactocentric distance R_{gc}. The metal-poor subpopulation ([Fe/H] < -1) is shown in solid dots, the metal-richer subpopulation ([Fe/H] > -1) as open symbols. For $R_{gc} \gtrsim 4$ kpc (solid line), a simple power-law dependence $\phi \sim R_{gc}^{-3.5}$ matches the spatial structure well, while for the inner bulge region, ϕ flattens off to something closer to an R^{-2} dependence. Notice that the metal-richer distribution falls off steeply for $R_{gc} \gtrsim 10$ kpc. This plot implicitly (and wrongly!) assumes a spherically symmetric space distribution, which smooths over any more detailed structure; see the discussion below

globular clusters by at least 100 to 1, the clusters are certainly the easiest halo objects to find). But we can see as well from Fig. 1.2 that the whole system is a *centrally concentrated* one, with the spatial density ϕ (number of clusters per unit volume in space) varying as $\phi \sim R_{gc}^{-3.5}$ over most of the halo (Fig. 1.3). Unfortunately, one immediate problem this leaves us is that more than half of our globular clusters can be studied only by peering in toward the Galactic center through the heavy obscuration of dust clouds in the foreground of the Galactic disk.[1] Until recent years our knowledge of

[1] For this reason, the YZ plane was used for the previous figure to display the large-scale space distribution. Our line of sight to most of the clusters is roughly parallel to the $X-$axis and thus any distance measurement error on our part will skew the estimated value of X much more than Y or Z. Of all possible projections, the YZ plane is therefore the most nearly "error-free" one.

these heavily reddened clusters remained surprisingly poor, and even today there are still a few clusters with exceptionally high reddenings embedded deep in the Galactic bulge about which we know almost nothing (see the listings in Harris 1996a).

However, progress over the years has been steady and substantial: compare the two graphs in Fig. 1.4. One (from the data of Shapley 1918) is the very first 'outside view' of the Milky Way GCS ever achieved, and the one used by Shapley to estimate the centroid of the system and thus – again for the first time – to determine the distance from the Sun to the Galactic center. The second graph shows us exactly the same plot with the most modern measurements. The data have improved dramatically over the intervening 80 years in three major ways: (1) The sample size of known clusters is now almost twice as large as Shapley's list. (2) Shapley's data took no account of reddening, since the presence and effect of interstellar dust was unknown then; the result was to overestimate the distances for most clusters and thus to elongate their whole distribution along the line of sight (roughly, the $X-$axis). (3) The fundamental distance scale used by Shapley – essentially, the luminosity of the RR Lyraes or the tip of the red-giant branch – was about one magnitude brighter than the value adopted today; again, the result was to overestimate distances for almost all clusters. Nevertheless, this simple diagram represented a breakthrough in the study of Galactic structure; armed with it, Shapley boldly argued both that the Sun was far from the center of the Milky Way, and that our Galaxy was much larger than had been previously thought.

The foreground reddening of any given cluster comes almost totally from dust clouds in the Galactic disk rather than the bulge or halo, and so reddening correlates strongly with Galactic latitude (Fig. 1.5). The basic cosecant-law dependence of E_{B-V} shows how very much more difficult it is to study objects at low latitudes. Even worse, such objects are also often afflicted with severe contamination by field stars and by differential (patchy) reddening. The equations for the reddening lines in Fig. 1.5 are:

Northern Galactic hemisphere ($b > 0$): $E_{B-V} = 0.060\ (\csc|b| - 1)$
Southern Galactic hemisphere ($b < 0$): $E_{B-V} = 0.045\ (\csc|b| - 1)$

Individual globular clusters have been known for at least two centuries. Is our census of them complete, or are we still missing some? This question has been asked many times, and attempted answers have differed quite a bit (e.g. Racine & Harris 1989; Arp 1965; Sharov 1976; Oort 1977; Barbuy et al. 1998). They are luminous objects, and easily found anywhere in the Galaxy as long as they are not either (a) *extremely* obscured by dust, or (b) too small and distant to have been picked up from existing all-sky surveys. Discoveries of faint, distant clusters at high latitude continue to happen occasionally as lucky accidents, but are now rare (just five new ones have been added over the last 20 years: AM-1 [Lauberts 1976; Madore & Arp 1979], Eridanus [Cesarsky et al. 1977], E3 [Lauberts 1976], Pyxis [Irwin et al. 1995], and IC

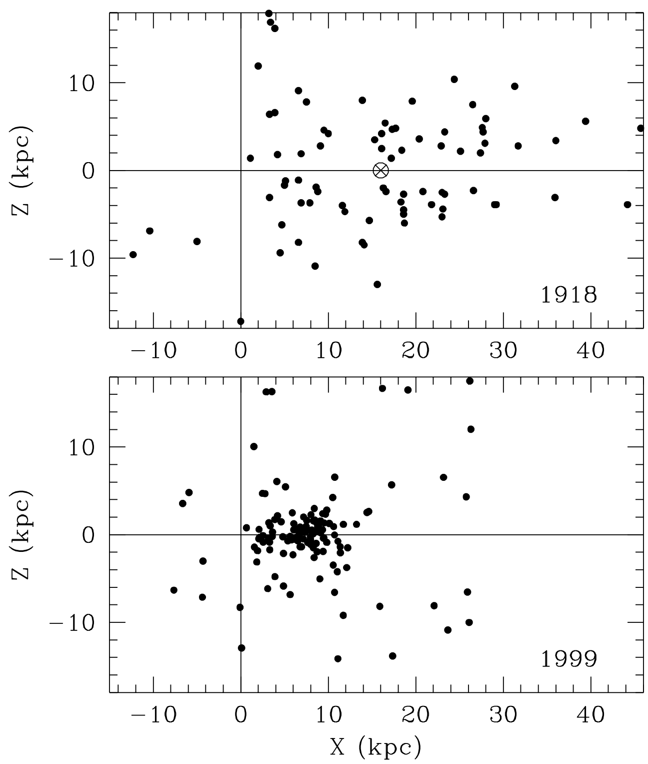

Fig. 1.4. *Upper panel:* The spatial distribution of the Milky Way clusters as measured by Shapley (1918). The Sun is at $(0,0)$ in this graph, and Shapley's estimated location of the Galactic center is marked at $(16,0)$. *Lower panel:* The spatial distribution in the same plane, according to the best data available today. The tight grouping of clusters near the Galactic center (now at $(8,0)$), and the underlying symmetry of the system, are now much more obvious

1257 [Harris et al. 1997a]). Recognizing the strong latitude effect that we see in Fig. 1.5, we might make a sensible estimate of missing heavily reddened clusters by using Fig. 1.6. The number of known clusters per unit latitude angle b rises exponentially to lower latitude, quite accurately as $n \sim e^{-|b|/14^\circ}$ for $2^\circ \lesssim b \lesssim 40^\circ$. It is only the first bin ($|b| < 2^\circ$) where incompleteness appears to be important; ~ 10 additional clusters would be needed there to bring the known sample back up to the curve.

Combining these arguments, I estimate that the total population of globular clusters in the Milky Way is $N = 160 \pm 10$, and that the existing sample is now likely to be more than 90% complete.

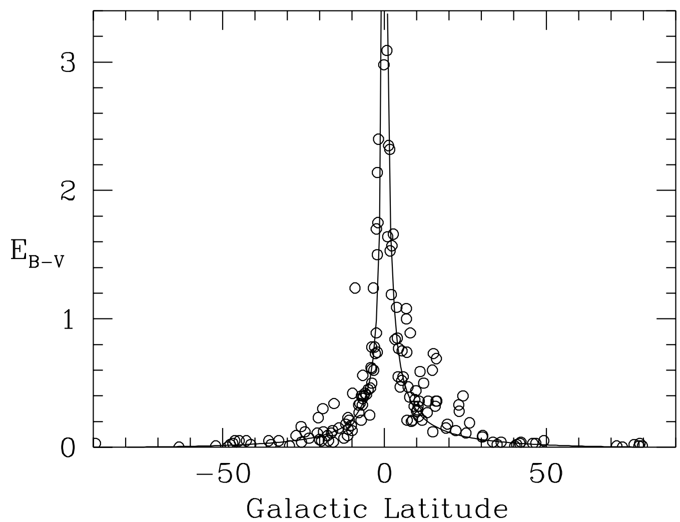

Fig. 1.5. The foreground reddening of globular clusters E_{B-V} plotted against Galactic latitude b (in degrees). The equations for the cosecant lines are given in the text

1.1.2 The Metallicity Distribution

The huge range in *heavy-element abundance* or *metallicity* among globular clusters became evident to spectroscopists half a century ago, when it was found that the spectral lines of the stars in most clusters were remarkably weak, resembling those of field subdwarf stars in the Solar neighborhood (e.g., Mayall 1946; Baum 1952; Roman 1952). Morgan (1956) and Baade at the landmark Vatican conference (1958) suggested that their compositions might be connected with Galactocentric location R_{gc} or Z. These ideas culminated in the classic work of Kinman (1959a,b), who systematically investigated the correlations among composition, location, and kinematics of subsamples within the GCS. By the beginning of the 1960's, these pioneering studies had been used to develop a prevailing view in which (a) the GCS possessed a *metallicity gradient*, with the higher-metallicity clusters residing only in the inner bulge regions and the average metallicity of the system declining steadily outwards; (b) the metal-poor clusters were a dynamically 'hot' system, with large random space motions and little systemic rotation; (c) the metal-richer clusters formed a 'cooler' subsystem with significant overall rotation and lower random motion.

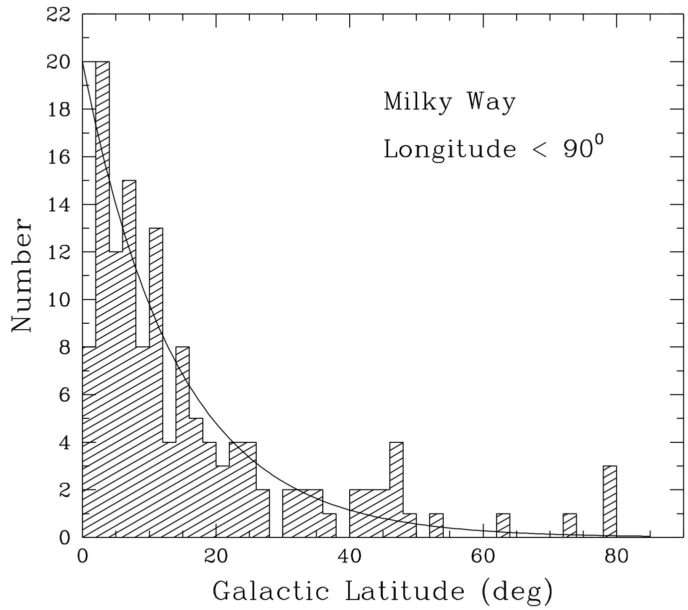

Fig. 1.6. Number of globular clusters as a function of Galactic latitude $|b|$, plotted in $2°$ bins; only the clusters within $\pm 90°$ longitude of the Galactic center are included. An exponential rise toward lower latitude, with an $e-$folding height of $14°$, is shown as the solid line

All of this evidence was thought to fit rather well into a picture for the formation of the Galaxy that was laid out by Eggen, Lynden-Bell, & Sandage (1962 [ELS]). In their model, the first stars to form in the protogalactic cloud were metal-poor and on chaotic, plunging orbits; as star formation continued, the remaining gas was gradually enriched, and as it collapsed inward and spun up, subsystems could form which were more and more disk-like. The timescale for all of this to take place could have been no shorter than the freefall time of the protogalactic cloud (a few 10^8 y), but might have been significantly longer depending on the degree of pressure support during the collapse. If pressure support was important, then a clear metallicity gradient should have been left behind, with cluster age correlated nicely with its chemical composition. The rough age calibrations of the globular clusters that were possible at the time (e.g., Sandage 1970) could not distinguish clearly between these alternatives, but were consistent with the view that the initial collapse was rapid.

This appealing model did not last – at least, not in its original simplicity. With steady improvements in the database, new features of the GCS emerged. One of the most important of these is the *bimodality of the cluster metallicity distribution*, shown in Fig. 1.7. Two rather distinct metallicity groups clearly exist, and it is immediately clear that the simple monolithic-collapse model for the formation of the GCS will not be adequate. To avoid prejudicing our view of these two subgroups as belonging to the Galactic halo, the disk, the bulge, or something else, I will simply refer to them as the *metal-poor component* (MPC) and the *metal-rich component* (MRC). In Fig. 1.7, the MPC has a fitted centroid at [Fe/H] $= -1.6$ and a dispersion $\sigma = 0.30$ dex, while the MRC has a centroid at [Fe/H] $= -0.6$ and dispersion $\sigma = 0.2$. The dividing line between them I will adopt, somewhat arbitrarily, at [Fe/H] $= -0.95$ (see the next section below).[2]

The distribution of [Fe/H] with location is shown in Fig. 1.8. Clearly, the dominant feature of this diagram is the *scatter* in metallicity at any radius R_{gc}. Smooth, pressure-supported collapse models are unlikely to produce a result like this. But can we see any traces at all of a metallicity gradient in which progressive enrichment occurred? For the moment, we will ignore the half-dozen remote objects with $R_{gc} > 50$ kpc (these "outermost-halo" clusters probably need to be treated separately, for additional reasons that we will see below). For the inner halo, a *small* net metallicity gradient is rather definitely present amidst the dominant scatter. Specifically, within both the MPC and MRC systems, we find Δ[Fe/H]/ $\Delta \log R_{gc} = -0.30$ for the restricted region $R_{gc} \lesssim 10$ kpc; that is, the heavy-element abundance scales as $(Z/Z_{\odot}) \sim R^{-0.3}$. At larger R_{gc}, no detectable gradient appears.

[2] Note that the [Fe/H] values used throughout my lectures are ones on the "Zinn-West" (ZW) metallicity scale, the most frequently employed system through the 1980's and 1990's. The large catalog of cluster abundances by Zinn & West (1984) and Zinn (1985) was assembled from a variety of abundance indicators including stellar spectroscopy, color-magnitude diagrams, and integrated colors and spectra. These were calibrated through high-dispersion stellar spectroscopy of a small number of clusters obtained in the pre-CCD era, mainly from the photographic spectra of Cohen (see Frogel et al. 1983 for a compilation). Since then, a much larger body of spectroscopic data has been built up and averaged into the ZW list, leading to a somewhat heterogeneous database (for example, see the [Fe/H] sources listed in the Harris 1996a catalog, which are on the ZW scale). More recently, comprehensive evidence has been assembled by Carretta & Gratton (1997) that the ZW [Fe/H] scale is nonlinear relative to contemporary high resolution spectroscopy, even though the older abundance indicators may still provide the correct *ranking* of relative metallicity for clusters. The problem is also discussed at length by Rutledge et al. (1997). Over the range containing most of the Milky Way clusters ([Fe/H] $\lesssim -0.8$) the scale discrepancies are not large (typically ± 0.2 dex at worst). But at the high$-$[Fe/H] end the disagreement becomes progressively worse, with the ZW scale overestimating the true [Fe/H] by ~ 0.5 dex at near-solar true metallicity. At time of writing these chapters, a completely homogeneous metallicity list based on the Carretta-Gratton scale has not yet been constructed. Lastly, it is worth emphasizing that the quoted metallicities for globular clusters are almost always based on spectral features of the highly evolved red giant stars. Eventually, we would like to base [Fe/H] on the (much fainter) unevolved stars.

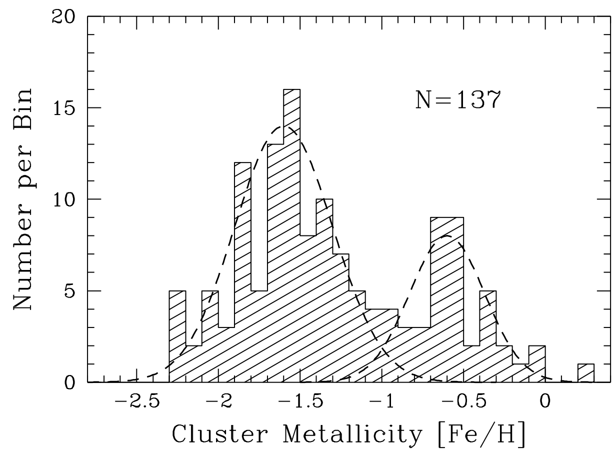

Fig. 1.7. Metallicity distribution for 137 Milky Way globular clusters with measured [Fe/H] values. The metallicities are on the Zinn-West (1984) scale, as listed in the current compilation of Harris (1996a). The bimodal nature of the histogram is shown by the two Gaussian curves whose parameters are described in the text

These features – the large scatter and modest inner-halo mean gradient – have been taken to indicate that the inner halo retains a trace of the classic monolithic rapid collapse, while the outer halo is dominated by chaotic formation and later accretion. They also helped stimulate a very different paradigm for the early evolution of the Milky Way, laid out in the papers of Searle (1977) and Searle & Zinn (1978 [SZ]). Unlike ELS, they proposed that the protoGalaxy was in a clumpy, chaotic, and non-equilibrium state in which the halo-star (and globular cluster) formation period could have lasted over many Gigayears. An additional key piece of evidence for their view was to be found in the connections among the *horizontal-branch morphologies* of the globular clusters, their locations in the halo, and their metallicities (Fig. 1.9). They noted that for the inner-halo clusters ($R_{gc} \lesssim R_0$), there was generally a close correlation between HB type and metallicity, as if all these clusters were the same age and HB morphology was determined only by metallicity. (More precisely, the same type of correlation would be generated if there were a one-to-one relation between cluster metallicity and age; i.e. if metallicity determined both age and HB type together. However, Lee et al. (1994) argue from isochrone models that the inner-halo correlation is nearly what we would expect for a single-age sequence differing only in metallicity.) In general, we can state that the morphology of the CMD is determined by several “parameters” which label the physical characteristics of the stars in

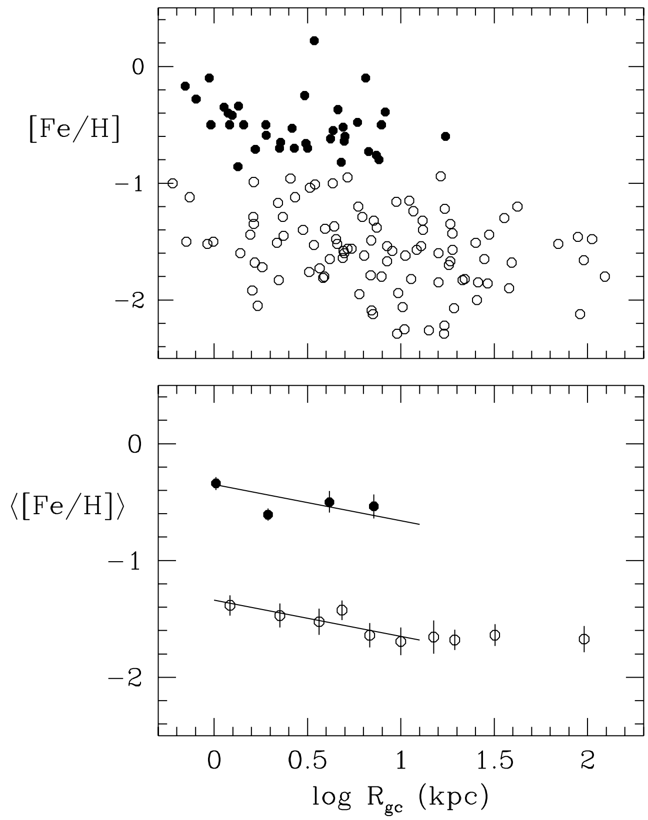

Fig. 1.8. [Fe/H] plotted against Galactocentric distance R_{gc}. *Upper panel*: Individual clusters are plotted, with MRC objects as solid symbols and MPC as open symbols. *Lower panel*: Mean [Fe/H] values for radial bins. Both MRC and MPC subsystems exhibit a slight gradient Δ[Fe/H]/ $\Delta \log R_{gc} = -0.30$ for $R_{gc} \lesssim 10$ kpc, as shown by the solid lines. For the more distant parts of the halo, no detectable mean gradient exists

the cluster. The *first parameter* which most strongly controls the distribution of stars in the CMD is commonly regarded to be metallicity, i.e. the overall heavy-element abundance. But quantities such as the HB morphology or the color and steepness of the giant branch do not correlate uniquely with only the metallicity, so more parameters must come into play. Which of these is most important is not known. At various times, plausible cases have been made that the dominant *second parameter* might be cluster age, helium abundance, CNO-group elements, or other factors such as mass loss or internal stellar rotation.

By contrast, for the intermediate- and outer-halo clusters the correlation between [Fe/H] and HB type becomes increasingly scattered, indicating that other parameters are affecting HB morphology just as strongly as metallicity. The interpretation offered by SZ was that the principal "second parameter" is age, in the sense that the *range in ages* is much larger for the outer-halo clusters. In addition, the progressive shift toward redder HBs at larger Galactocentric distance (toward the left in Fig. 1.9) would indicate a trend toward lower mean age in this interpretive picture.

From the three main pieces of evidence (a) the large scatter in [Fe/H] at any location in the halo, (b) the small net gradient in mean [Fe/H], and (c) the weaker correlation between HB type and metallicity at increasing R_{gc}, SZ concluded that the entire halo could not have formed in a pressure-supported monolithic collapse. Though the inner halo could have formed with some degree of the ELS-style formation, the outer halo was dominated by chaotic formation and even accretion of fragments from outside. They suggested that the likely formation sites of globular clusters were within large individual gas clouds (to be thought of as protogalactic 'fragments'), within which the compositions of the clusters were determined by very local enrichment processes rather than global ones spanning the whole protogalactic potential well. Although a large age range is not *necessary* in this scheme (particularly if other factors than age turn out to drive HB morphology strongly), a significant age range would be much easier to understand in the SZ scenario, and it opened up a wide new range of possibilities for the way halos are constructed. We will return to further developments of this picture in later sections. For the moment, we will note only that, over the next two decades, much of the work on increasingly accurate age determination and composition analysis for globular clusters all over the Galactic halo was driven by the desire to explore this roughed-out model of 'piecemeal' galaxy formation.

1.1.3 The Metal-Rich Population: Disk or Bulge?

Early suggestions of distinct components in the metallicity distribution were made by, e.g., Marsakov & Suchkov (1976) and Harris & Canterna (1979), but it was the landmark paper of Zinn (1985) which firmly identified two distinct subpopulations *and* showed that these two groups of clusters also had distinct kinematics and spatial distributions. In effect, *it was no longer possible to talk about the GCS as a single stellar population.* Our next task is, again, something of a historically based one: using the most recent data, we will step through a classic series of questions about the nature of the MRC and MPC.

The spatial distributions of the MPC and MRC are shown in Fig. 1.10. Obviously, the MRC clusters form a subsystem with a much smaller scale size.

Since the work of Zinn (1985) and Armandroff (1989), the MRC has conventionally been referred to as a "disk cluster" system, with the suggestion

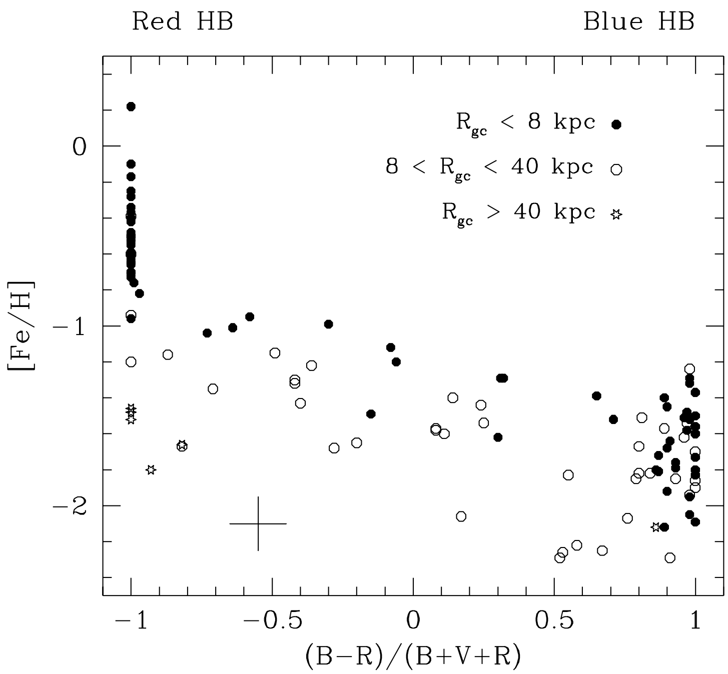

Fig. 1.9. Metallicity versus horizontal branch type for globular clusters. The HB ratio (B-R)/(B+V+R) (Lee et al. 1994) is equal to −1 for clusters with purely red HBs, increasing to +1 for purely blue HBs. A typical measurement uncertainty for each point is shown at lower left. Data are taken from the catalog of Harris (1996a)

that these clusters belonged spatially and kinematically to the thick disk. This question has been re-investigated by Minniti (1995) and Côté (1999), who make the case that they are better associated with the Galactic bulge. A key observation is the fact that the relative number of the two types of clusters, N_{MRC}/N_{MPC}, *rises steadily inward to the Galactic center*, in much the same way as the bulge-to-halo-star ratio changes inward, whereas in a true "thick-disk" population this ratio should die out to near-zero for $R_{gc} \lesssim 2$ kpc.

The MRC space distribution is also not just a more compact version of the MPC; rather, it appears to be genuinely flattened toward the plane. A useful diagnostic of the subsystem shape is to employ the angles (ω, θ) defining the cluster location on the sky relative to the Galactic center (see Zinn 1985 and Fig. 1.11). Consider a vector from the Galactic center to the cluster as

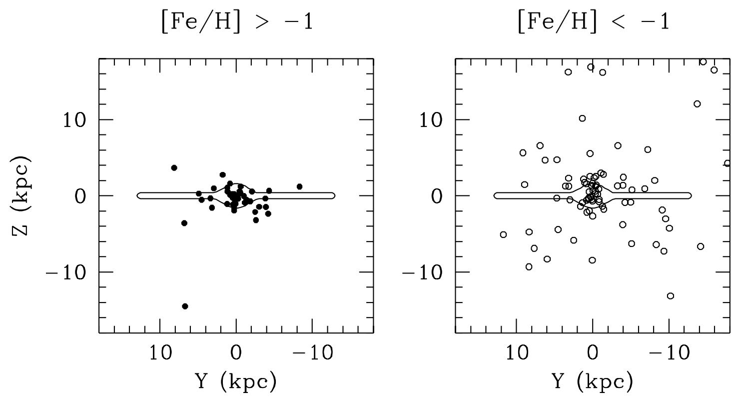

Fig. 1.10. Spatial distribution projected on the YZ plane for the metal-rich clusters (left panel) with [Fe/H]> -0.95, and the metal-poor clusters (right panel) with [Fe/H]< -0.95. In the left panel, the most extreme outlying point is Palomar 12, a "transition" object between halo and disk

seen projected on the sky: the angular length of the vector is ω, while the orientation angle between ω and the Galactic plane is θ:

$$\cos\omega = \cos b \cos\ell\,, \quad \tan\theta = \tan b \csc\ell\ . \tag{1.2}$$

In Fig. 1.12, the (ω, θ) point distributions are shown separately for the MRC and MPC subsystems. Both graphs have more points at smaller ω, as is expected for any population which is concentrated toward the Galactic center. However, any spherically symmetric population will be uniformly distributed in the azimuthal angle θ, whereas a flattened (disk or bulge) population will be biased toward small values of θ. The comparison test must also recognize the probable incompleteness of each sample at low latitude: for $|b| \lesssim 3°$, the foreground absorption becomes extremely large, and fewer objects appear below that line in either diagram.

A marked difference between the two samples emerges (Table 1.1) if we simply compare the mean $\langle\theta\rangle$ for clusters within 20° of the center (for which the effects of reddening should be closely similar on each population). A population of objects which has a spherical spatial distribution *and is unaffected by latitude incompleteness* would have $\langle\theta\rangle = 45°$, whereas sample incompleteness at low b would bias the mean $\langle\theta\rangle$ to higher values. Indeed, the MPC value $\langle\theta\rangle = 57°$ is consistent with that hypothesis – that is, that low-latitude clusters are missing from the sample because of their extremely high reddenings. However, the MRC value $\langle\theta\rangle = 40°$ – which must be affected by the *same* low-latitude incompleteness – can then result only if it

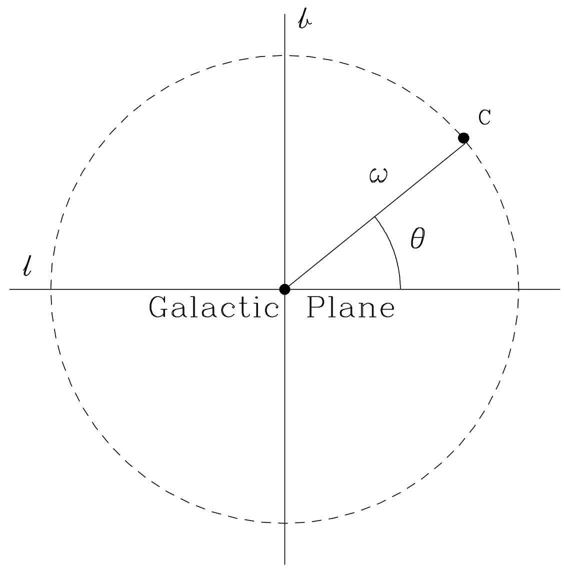

Fig. 1.11. Definition of the angles ω, θ given in the text: the page represents the plane of the sky, centered on the Galactic center. The Galactic longitude and latitude axes (ℓ, b) are drawn in. The distance from the Galactic center to the cluster C subtends angle ω, while θ is its orientation angle to the Galactic plane

belongs to an intrinsically flattened distribution. A Kolmogorov-Smirnov test on the $\theta-$distribution confirms that these samples are different at the $\sim 93\%$ confidence level, in the sense that the MRC is more flattened.

Another way to define the same result is to compare the linear coordinates Z, Y, and $\sqrt{X^2+Y^2}$ (Table 1.1). The relevant ratios Z/Y and $Z/\sqrt{X^2+Y^2}$ are half as large for the MRC as for the MPC, again indicating a greater flattening to the plane.

Table 1.1. Spatial flattening parameters for bulge clusters

	$\langle\theta\rangle$ (deg)	$\langle\|Z\|\rangle/\langle\|Y\|\rangle$	$\langle\|Z\|\rangle/\langle\sqrt{X^2+Y^2}\rangle$
MRC ($\omega < 20^o$)	40.0 ± 4.8	0.81 ± 0.20	0.37 ± 0.08
MPC ($\omega < 20^o$)	56.8 ± 4.5	1.63 ± 0.39	0.68 ± 0.13

Our tentative conclusion from these arguments is that the *inner* MRC – the clusters within $\omega \sim 20°$ or 3 kpc of the Galactic center – outline something best resembling a *flattened bulge population*. Kinematical evidence will be added in Section 4.

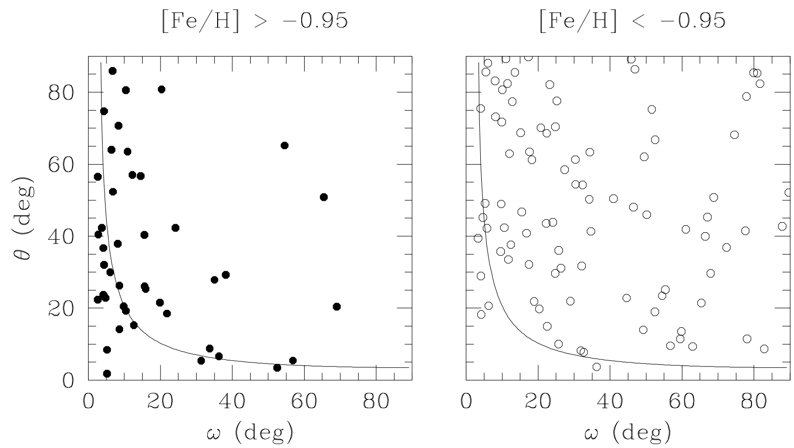

Fig. 1.12. Spatial distribution diagnostics for the MRC (left panel) and MPC (right panel) clusters. Here ω is the angle between the Galactic center and the cluster as seen on the sky, and θ is the angle between the Galactic plane and the line joining the Galactic center and the cluster. A line of constant Galactic latitude ($b = 3.5$ degrees) is shown as the curved line in each figure. Below this line, the foreground reddening becomes large and incompleteness in both samples is expected

1.1.4 The Distance to the Galactic Center

As noted above, Shapley (1918) laid out the definitive demonstration that the Sun is far from the center of the Milky Way. His first estimate of the distance to the Galactic center was $R_0 = 16$ kpc, only a factor of two different from today's best estimates (compare the history of the Hubble constant over the same interval!). In the absence of sample selection effects and measurement biases, Shapley's hypothesis can be written simply as $R_0 = \langle X \rangle$ where the mean $X-$coordinate is taken over the entire globular cluster population (indeed, the same relation can be stated for any population of objects centered at the same place, such as RR Lyraes, Miras, or other standard candles). But of course the sample mean $\langle X \rangle$ is biased especially by incompleteness and nonuniformity at low latitude, as well as distortions in converting distance modulus $(m - M)_V$ to linear distance X: systematic errors will result if the reddening is estimated incorrectly or if the distance-scale calibration for $M_V(HB)$ is wrong. Even the random errors of measurement in distance modulus convert to asymmetric error bars in X and thus a systematic bias in $\langle X \rangle$. One could minimize these errors by simply ignoring the "difficult" clusters at low latitude and using only low-reddening clusters at high latitude. However, there are not that many high-halo clusters ($N \sim 50$), and they are widely spread through the halo, leaving an uncomfortably large and irreducible uncertainty of $\sim \pm 1.5$ kpc in the centroid position $\langle X \rangle$ (see, e.g., Harris 1976 for a thorough discussion).

A better method, outlined by Racine & Harris (1989), is to use the *inner* clusters and to turn their large and different reddenings into a partial advantage. The basic idea is that, to first order, the great majority of the clusters we see near the direction of the Galactic center are *at the same true distance* R_0 – that is, they are *in* the Galactic bulge, give or take a kiloparsec or so – despite the fact that they may have wildly different *apparent* distance moduli.[3] This conclusion is guaranteed by the strong central concentration of the GCS (Fig. 1.2) and can be quickly verified by simulations (see Racine & Harris). For the inner clusters, we can then write $d \simeq R_0$ for essentially all of them, and thus

$$(m-M)_V \equiv (m-M)_0 + A_V \simeq const + R \cdot E_{B-V} \tag{1.3}$$

where $R \simeq 3.1$ is the adopted ratio of total to selective absorption. Now since the horizontal-branch magnitude V_{HB} is a good indicator of the cluster apparent distance modulus, varying only weakly with metallicity, a simple graph of V_{HB} against reddening for the inner globular clusters should reveal a straight-line relation with a (known) slope equal to R:

$$V_{HB} \simeq M_V(HB) + 5\log(R_0/10\mathrm{pc}) + R \cdot E_{B-V} \tag{1.4}$$

The observed correlation is shown in Fig. 1.13. Here, the "component" of V_{HB} projected onto the $X-$axis, namely $V_{HB} + 5 \log(\cos \omega)$, is plotted against reddening. As we expected, it resembles a distribution of objects which are all at the same *true* distance d (with some scatter, of course) but with different amounts of foreground reddening. There are only 4 or 5 obvious outliers which are clearly well in front of or behind the Galactic bulge. The quantity we are interested in is the *intercept* of the relation, i.e. the value at zero reddening. This intercept represents the distance modulus of an *unreddened* cluster which is *directly at the Galactic center*.

[3] It is important to realize that the clusters nearest the Galactic center, because of their low Galactic latitude, are reddened *both* by local dust clouds in the Galactic disk near the Sun *and* by dust in the Galactic bulge itself. In most cases the contribution from the nearby dust clouds is the dominant one. Thus, the true distances of the clusters are almost uncorrelated with foreground reddening (see also Barbuy et al. 1998 for an explicit demonstration). Clusters on the far side of the Galactic center are readily visible in the normal optical bandpasses unless their latitudes are $\lesssim 1°$ or $2°$.

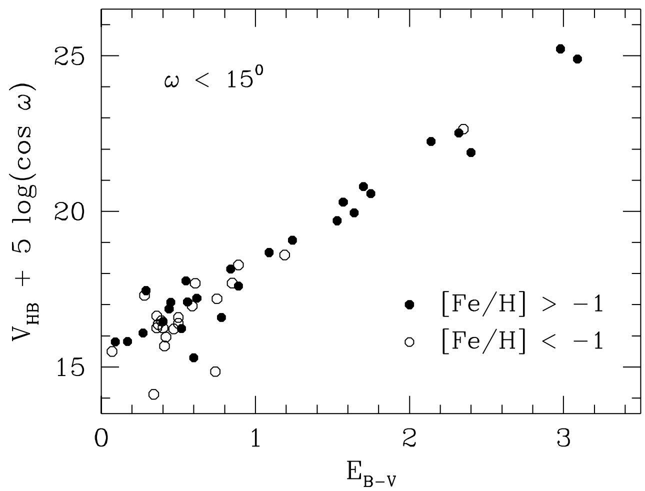

Fig. 1.13. Apparent magnitude of the horizontal branch plotted against reddening, for all globular clusters within $\omega = 15°$ of the Galactic center. MRC and MPC clusters are in solid and open symbols

We can refine things a bit more by taking out the known second-order dependence of A_V on E_{B-V}, as well as the dependence of V_{HB} on metallicity. Following Racine & Harris, we define a linearized HB level as

$$V_{HB}^C \ = \ V_{HB} + 5\log(cos\,\omega) - 0.05\,E_{B-V}^2 - 0.15\,([\mathrm{Fe/H}]\ +\ 2.0) \qquad (1.5)$$

The correlation of V_{HB}^C with E_{B-V} is shown in Fig. 1.14. Ignoring the 5 most deviant points at low reddening, we derive a best-fit line

$$V_{HB}^C \ = \ (15.103 \pm 0.123) + (2.946 \pm 0.127)\,E_{B-V} \qquad (1.6)$$

with a remaining r.m.s. scatter of ± 0.40 in distance modulus about the mean line. The slope of the line $\Delta V/\Delta E_{B-V} \sim 3$ is just what it should be if it is determined principally by reddening differences *that are uncorrelated with true distance.* The intercept is converted into the distance modulus of the Galactic center by subtracting our distance scale calibration $M_V(HB) = 0.50$ at [Fe/H]$= -2.0$. We must also remove a small geometric bias of 0.05 ± 0.03 (Racine & Harris) to take account of the fact that our line-of-sight cone defined by $\omega < 15°$ has larger volume (and thus proportionally more clusters) beyond the Galactic center than in front of it. The error budget will also include $\Delta(m - M) \sim \pm 0.1$ (internal) due to uncertainty in the reddening law, and (pessimistically, perhaps) a $\pm 0.2-$mag external uncertainty

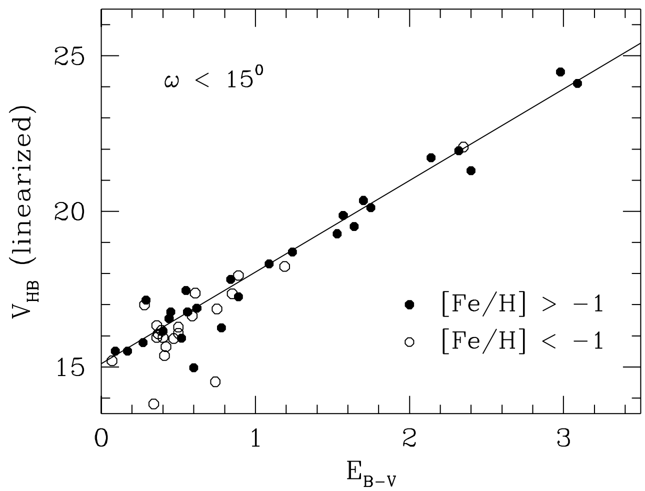

Fig. 1.14. Apparent magnitude of the horizontal branch plotted against reddening, after projection onto the X−axis and correction for second-order reddening and metallicity terms. The equation for the best-fit line shown is given in the text; it has a slope $R \sim 3$ determined by foreground reddening. The intercept marks the distance modulus to the Galactic center

in the distance scale zeropoint. In total, our derived distance modulus is $(m-M)_0(GC) = 14.55 \pm 0.16(\text{int}) \ \pm 0.2(\text{ext})$, or

$$R_0 = 8.14 \text{ kpc } \ \pm 0.61 \text{ kpc(int) } \ \pm 0.77 \text{ kpc(ext)} \tag{1.7}$$

It is interesting that the dominant source of uncertainty is in the luminosity of our fundamental standard candle, the RR Lyrae stars. By comparison, the intrinsic cluster-to-cluster scatter of distances in the bulge creates only a $\pm 0.35-$kpc uncertainty in R_0.

This completes our review of the spatial distribution of the GCS, and the definition of its two major subpopulations. However, before we go on to discuss the kinematics and dynamics of the system, we need to take a more careful look at justifying our fundamental distance scale. That will be the task for the next Section.

1.2 THE DISTANCE SCALE

The researches of many commentators have already thrown much darkness on this subject, and it is probable that, if they continue, we shall soon know nothing at all about it.

Mark Twain

About 40 years ago, there was a highly popular quiz show on American television called "I've Got a Secret". On each show, three contestants would come in and all pretend to be the same person, invariably someone with an unusual or little-known occupation or accomplishment. Only one of the three was the real person. The four regular panellists on the show would have to ask them clever questions, and by judging how realistic the answers sounded, decide which ones were the imposters. The entertainment, of course, was in how inventive the contestants could be to fool the panellists for as long as possible. At the end of the show, the moderator would stop the process and ask the real contestant to stand up, after which everything was revealed.

The metaphor for this section is, therefore, "Will the real distance scale please stand up?" In our case, however, the game has now gone on for a century, and there is no moderator. For globular clusters and Population II stars, there are several routes to calibrating distances. These routes do not agree with one another; and the implications for such things as the cluster ages and the cosmological distance scale are serious. It is a surprisingly hard problem to solve, and at least some of the methods we are using must be wrong. But which ones, and how?

The time-honored approach to calibrating globular cluster distances is to measure some identifiable sequence of stars in the cluster CMD, and then to establish the luminosities of these same types of stars in the Solar neighborhood by trigonometric parallax. The three most obvious such sequences (see the Appendix) are:

- *The horizontal branch*, or RR Lyrae stars: In the V band, these produce a sharp, nearly level and thus almost ideal sequence in the CMD. The problem is in the comparison objects: field RR Lyrae variables are rare and uncomfortably distant, and thus present difficult targets for parallax programs. There is also the nagging worry that the field halo stars may be astrophysically different (in age or detailed chemical composition) from those in clusters, and the HB luminosity depends on many factors since it represents a rather advanced evolutionary stage. The HB absolute magnitude almost certainly depends weakly on metallicity. It is usual to parametrize this effect simply as $M_V(HB) = \alpha$ [Fe/H] $+\beta$, where (α, β) are to be determined from observations – and, we hope, with some guidance from theory.[4]

[4] As noted in the previous Section, I define V_{HB} as the mean magnitude of the the horizontal-branch stars without adjustment. Some other authors correct V_{HB} to the slightly fainter level of the "zero-age" unevolved ZAHB.

- *The unevolved main sequence (ZAMS)*: modern photometric tools can now establish highly precise main sequences for any cluster in the Galaxy not affected by differential reddening or severe crowding. As above, the problem is with the comparison objects, which are the unevolved halo stars or "subdwarfs" in the Solar neighborhood. Not many are near enough to have genuinely reliable parallaxes even with the new *Hipparcos* measurements. This is particularly true for the lowest-metallicity ones which are the most relevant to the halo globular clusters; and most of them do not have accurate and detailed chemical compositions determined from high-dispersion spectroscopy.
- *The white dwarf sequence*: this faintest of all stellar sequences has now come within reach from *HST* photometry for a few clusters. Since its position in the CMD is driven by different stellar physics than is the main sequence or HB, it can provide a uniquely different check on the distance scale. Although such stars are common, they are so intrinsically faint that they must be *very* close to the Sun to be identified and measured, and thus only a few comparison field-halo white dwarfs have well established distances.

These classic approaches each have distinct advantages and problems, and other ways have been developed to complement them. In the sections below, I provide a list of the current methods which seem to me to be competitive ones, along with their results. Before we plunge into the details, I stress that this whole subject area comprises a vast literature, and we can pretend to do nothing more here than to select recent highlights.

1.2.1 Statistical Parallax of Field Halo RR Lyraes

Both the globular clusters and the field RR Lyrae stars in the Galactic halo are too thinly scattered in space for almost any of them to lie within the distance range of direct trigonometric parallax. However, the radial velocities and proper motions of the field RR Lyraes can be used to solve for their luminosity through statistical parallax. In principle, the trend of luminosity with metallicity can also be obtained if we divide the sample up into metallicity groups.

An exhaustive analysis of the technique, employing ground-based velocities and Lick Observatory proper motions, is presented by Layden et al. (1996). They use data for a total of 162 "halo" (metal-poor) RR Lyraes and 51 "thick disk" (more metal-rich) stars in two separate solutions, with results as shown in Table 1.2. Recent solutions are also published by Fernley et al. (1998a), who use proper motions from the *Hipparcos* satellite program; and by Gould & Popowski (1998), who use a combination of Lick ground-based and *Hipparcos* proper motions. These studies are in excellent agreement with one another, and indicate as well that the metallicity dependence of $M_V(RR)$ is small. The statistical-parallax calibration traditionally gives lower-luminosity results than most other methods, but if there are problems in its assumptions that would systematically affect the results by more than its internal uncertainties, it is not yet clear what they might be. The discussion of Layden et al. should be referred to for a thorough analysis of the possibilities.

Table 1.2. Statistical parallax calibrations of field RR Lyrae stars

Region	$M_V(RR)$	[Fe/H]	Source
Halo	0.71 ± 0.12	-1.61	Layden et al.
Halo	0.77 ± 0.17	-1.66	Fernley et al.
Halo	0.77 ± 0.12	-1.60	Gould & Popowski
Disk	0.79 ± 0.30	-0.76	Layden et al.
Disk	0.69 ± 0.21	-0.85	Fernley et al.

1.2.2 Baade-Wesselink Method

This technique, which employs simultaneous radial velocity and photometric measurements during the RR Lyrae pulsation cycle, is discussed in more detail in this volume by Carney; here, I list only some of the most recent results. A synthesis of the data for 18 field RR Lyrae variables over a wide range of metallicity (Carney, Storm, & Jones 1992) gives

$$M_V(RR) \, = \, (0.16 \pm 0.03) \, [\mathrm{Fe/H}] + (1.02 \pm 0.03) \tag{1.8}$$

As Carney argues, the uncertainty in the *zeropoint* of this relation quoted above is only the internal uncertainty given the assumptions in the geometry of the method; the external uncertainty is potentially much larger. However, the *slope* is much more well determined and is one of the strongest aspects of the method if one has a sample of stars covering a wide metallicity range (see also Carney's lectures in this volume, and Fernley et al. 1998b for additional discussion of the slope α).

The Baade-Wesselink method can also be applied to RR Lyraes that are directly in globular clusters; although these are much fainter than the nearest field stars and thus more difficult to observe, at least this approach alleviates concerns about possible differences between field RR Lyraes and those in clusters. Recent published results for four clusters are listed in Table 1.3 (from Liu & Janes 1990; Cohen 1992; and Storm et al. 1994a,b). The third column of the table gives the measured $M_V(RR)$, while for comparison the fourth column gives the expected M_V from the field-star equation above. Within the uncertainties of either method, it is clear that the statistical parallax and Baade-Wesselink measurements are in reasonable agreement.

Table 1.3. Baade-Wesselink calibrations of RR Lyrae stars in four clusters

Cluster	[Fe/H]	$M_V(BW)$	M_V(eqn)
M92	-2.3	0.44, 0.64	0.65
M5	-1.3	0.60	0.81
M4	-1.2	0.80	0.83
47 Tuc	-0.76	0.71	0.90

1.2.3 Trigonometric Parallaxes of HB Stars

The *Hipparcos* catalog of trigonometric parallaxes provides several useful measurements of field HB stars for the first time (see Fernley et al. 1998a; Gratton 1998). One of these is RR Lyrae itself, for which $\pi = (4.38 \pm 0.59)$ mas, yielding $M_V(RR) = 0.78 \pm 0.29$ at [Fe/H] $= -1.39$. The red HB star HD 17072 (presumably a more metal-rich one than RR Lyrae) has a slightly better determined luminosity at $M_V(HB) = 0.97 \pm 0.15$. Finally, Gratton (1998) derives a parallax-weighted mean luminosity for $\sim$ 20 HB stars of $M_V(HB) = 0.69 \pm 0.10$ at a mean metallicity $\langle$[Fe/H]$\rangle = -1.41$, though of course the parallaxes for any individual HB star in this list are highly uncertain. At a given metallicity, these HB luminosities tend to sit $\sim 0.1 - 0.2$ mag higher than the ones from statistical parallax and Baade-Wesselink.

1.2.4 Astrometric Parallax

We turn next to distance calibration methods of other types, which can be used secondarily to establish $M_V(HB)$.

An ingenious method applying directly to clusters without the intermediate step of field stars, and without requiring any knowledge of their astrophysical properties, is that of "astrometric parallax": the internal motions of the stars within a cluster can be measured either through their radial velocity dispersion $\sigma(v_r)$, or through their dispersion in the projected radial and tangential proper motions $\sigma(\mu_r, \mu_\theta)$ relative to the cluster center. These three internal velocity components can be set equal through a simple scale factor involving the distance d,

$$\sigma(v_r) \; = \; const \, \cdot \, d \, \cdot \, \sigma(\mu) \tag{1.9}$$

and thus inverted to yield d, independent of other factors such as cluster metallicity and reddening. The two $\mu-$dispersions can also be used to model any radial anisotropy of the internal motions, and thus to adjust the scaling to $\sigma(v_r)$.

This method is in principle an attractive and powerful one, though the available measurements do not yet reach a level of precision for individual clusters that is sufficient to confirm or rule out other approaches definitively. A preliminary summary of the current results by Rees (1996) gives distances for five intermediate-metallicity clusters (M2, M4, M5, M13, M22, with a mean $\langle[\mathrm{Fe/H}]\rangle = -1.46$) equivalent to $M_V(HB) = 0.63 \pm 0.11$. For one low-metallicity cluster (M92, at [Fe/H] $= -2.3$), he finds $M_V(HB) = 0.31 \pm 0.32$. On average, these levels are $\sim 0.1 - 0.2$ mag brighter than the results from statistical parallax or Baade-Wesselink.

1.2.5 White Dwarf Sequences

Recently Renzini et al. (1996) have used deep *HST* photometry to establish the location of the WD sequence in the low-metallicity cluster NGC 6752 and to match it to five DA white dwarfs in the nearby field. The quality of the fit is remarkably tight even given the relatively small number of stars. The derived distance modulus corresponds to $M_V(HB) = 0.52 \pm 0.08$ at a cluster metallicity [Fe/H] $= -1.55$. The critical underlying assumption here is that the *mass* of the white dwarfs in the cluster – which is the most important determinant of the WD sequence luminosity – has the same canonical value $\simeq 0.6 M_\odot$ as the field DA's.

In a comparably deep photometric study of the nearby cluster M4, Richer et al. (1995) take the argument in the opposite direction: by using the heavily populated and well defined WD sequence along with a distance derived from main sequence fitting, they derive the WD mass, which turns out to be $\simeq 0.50 - 0.55 M_\odot$. A third deep white dwarf sequence has been measured for NGC 6397 by Cool et al. (1996), again with similar results, and HST-based results for other clusters are forthcoming. Thus at the present time, it appears that the fundamental distance scale from WDs is consistent with the range of numbers from the other approaches and deserves to be given significant weight. We can look forward, in a few years time, to a much more complete understanding of the relative WD vs. ZAMS distance scales and to a stronger contribution to the zeropoint calibration. Still deeper observations will, eventually, be able to find the faint-end termination of the WD sequence and place completely new observational limits on the cluster ages.

1.2.6 Field Subdwarf Parallaxes and Main Sequence Fitting

The technique which has generated the most vivid recent discussion (and controversy) centers on the matching of nearby halo main-sequence stars (subdwarfs) to cluster main sequences. It was widely expected that the *Hipparcos* project would, for the first time, supply a large number of high-quality trigonometric parallaxes for low-metallicity stars in the Solar neighborhood and would essentially solve the distance scale problem at a level which could claim to being definitive. Unfortunately, this hope has not been borne out.

The whole problem in the fitting procedure is essentially that any given collection of subdwarfs does not automatically give us a "sequence" which can then be matched immediately to a globular cluster. The individual subdwarfs all have different distances (and thus parallax uncertainties) and metallicities.

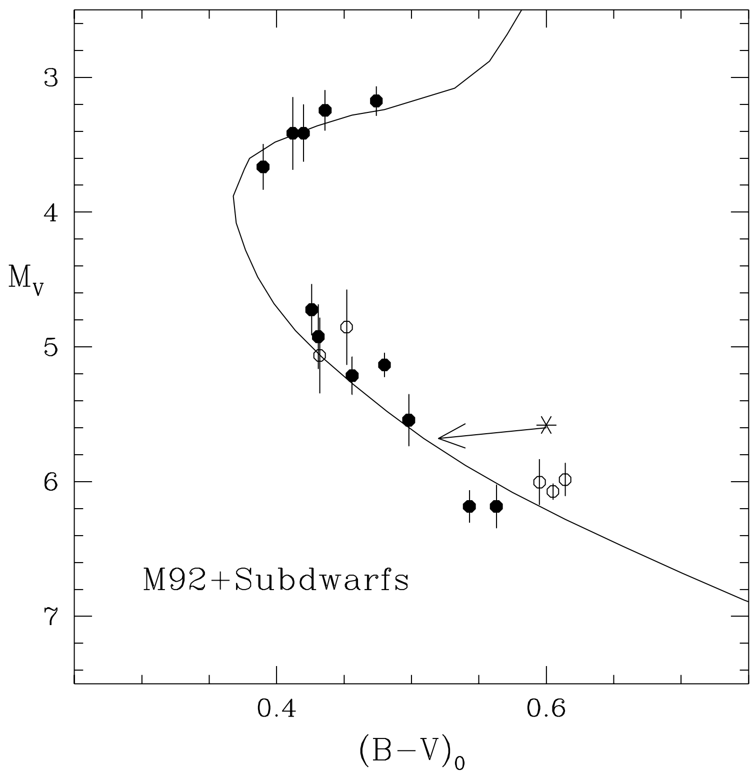

Fig. 1.15. An illustration of subdwarf fitting to a cluster main sequence. Nearby metal-poor subdwarfs (Pont et al. 1998), shown as the dots, are superimposed on the fiducial sequence for the metal-poor cluster M92 (Stetson & Harris 1988), for an assumed reddening $E(B-V) = 0.02$ and a distance modulus $(m-M)_V = 14.72$. The location of each star on this diagram must be adjusted to the color and luminosity it would have at the metallicity of M92 ([Fe/H] = −2.2). For a typical subdwarf at [Fe/H] ~ -1.6 (starred symbol), the size of the color and luminosity corrections is indicated by the arrow. The luminosity and color corrections follow the bias prescriptions in Pont et al. Known or suspected *binary* stars are plotted as open circles

All of them have to be relocated in the CMD back to the positions they would have *at the metallicity of the cluster*, and various biases may exist in the measured luminosities (see below). The more distant or low-latitude ones may even have small amounts of reddening, and the sample may also include undetected binaries. Thus before any fit to a given cluster can be done, a fiducial main sequence must be constructed out of a collection of subdwarfs which *by definition* is heterogeneous.

Figure 1.15 illustrates the procedure. The luminosity M_V of a given subdwarf, calculated directly from its raw trigonometric parallax and apparent magnitude (starred symbol in the figure), is adjusted by an amount ΔM_V for various sample bias corrections as described below. Next, the raw color index $(B-V)$ is adjusted by an amount $\Delta(B-V)$ to compensate for the difference in metallicity between subdwarf and cluster, and also for any reddening difference between the two. Usually $\Delta(B-V)$ is negative since most of the known subdwarfs are more metal-rich than most of the halo globular clusters, and the main sequence position becomes bluer at lower metallicity. The change of color with metallicity is normally calculated from theoretical isochrones; although this is the only point in the argument which is model dependent, it is generally regarded as reliable to ± 0.01 for the most commonly used indices such as $(B-V)$ or $(V-I)$ (the differential color shifts with metallicity are quite consistent in isochrones from different workers, even if the absolute positions may differ slightly).

The greatest concerns surround (a) the believed absolute accuracy of the published parallaxes, and (b) the degree to which bias corrections should be applied to the measured luminosities. These biases include, but are not limited to, the following effects:

- The Lutz-Kelker (1973) effect, which arises in parallax measurement of any sample of physically identical stars which are scattered at different distances. Since the volume of space sampled increases with distance, there will be more stars at a given π that were scattered inward by random measurement error from larger distances than outward from smaller distances. The deduced luminosity M_V of the stars therefore tends statistically to be too faint, by an amount which increases with the relative measurement uncertainty σ_π/π (see Hanson 1979 and Carretta et al. 1999 for a comprehensive discussion and prescriptions for the correction).
- The binary nature of some of the subdwarfs, which (if it lurks undetected) will bias the luminosities upward.
- The strong increase of σ_π with V magnitude (fainter stars are more difficult to measure with the same precision). This effect tends to remove intrinsically fainter stars from the sample, and also favors the accidental inclusion of binaries (which are brighter than single stars at the same parallax).
- The metallicity distribution of the known subdwarfs, which is asymmetric and biased toward the more common higher-metallicity (redder) stars. In any selected sample, accidental inclusion of a higher-metallicity star is thus more likely than a lower-metallicity one, which is equivalent to a mean sample luminosity that is too high at a given color.

It is evident that the various possible luminosity biases can act in opposite directions, and that a great deal of information about the subdwarf sample must be in hand to deal with them correctly. Four recent studies are representative of the current situation. Reid (1997) uses a sample of 18 subdwarfs with $\sigma_\pi/\pi < 0.12$ along with the Lutz-Kelker corrections and metallicity adjustments to derive new distances to five nearby clusters of low reddening. Gratton et al. (1997) use a different sample of 13 subdwarfs, again with $\sigma_\pi/\pi < 0.12$, and exert considerable effort to correct for the presence of binaries. They use Monte Carlo simulations to make further (small) corrections for parallax biases, and derive distances to nine nearby clusters. When plotted against metallicity, these define a mean sequence

$$M_V(HB) = (0.125 \pm 0.055)\,[\mathrm{Fe/H}] + (0.542 \pm 0.090) \tag{1.10}$$

which may be compared (for example) with the much fainter Baade-Wesselink sequence listed earlier. Pont et al. (1998) employ still another sample of 18 subdwarfs and subgiants with $\sigma_\pi/\pi \lesssim 0.15$ and do more Monte Carlo modelling to take into account several known bias effects simultaneously. They find that the net bias correction ΔM_V is small – nearly negligible for [Fe/H] ~ -1 and only $+0.06$ for [Fe/H] ~ -2. They derive a distance only to M92, the most metal-poor of the standard halo clusters, with a result only slightly lower than either Reid or Gratton et al. found. Lastly, a larger set of 56 subdwarfs drawn from the entire *Hipparcos* database is analyzed by Carretta et al. (1999), along with a comprehensive discussion of the bias corrections. Their results fall within the same range as the previous three papers.

Regardless of the details of the fitting procedure, the basic effect to be recognized is that the *Hipparcos* parallax measurements for the nearby subdwarfs tend to be a surprising ~ 3 milliarcseconds smaller than previous ground-based measurements gave. This difference then translates into brighter luminosities by typically $\Delta M_V \sim 0.2 - 0.3$ mag (see Gratton et al.). At the low metallicity end of the globular cluster scale ([Fe/H] $\simeq -2.2$, appropriate to M92 or M15), the *Hipparcos*-based analyses yield $M_V(HB) \simeq 0.3 \pm 0.1$, a level which is $\gtrsim 0.3$ mag brighter than (e.g.) from statistical parallax or Baade-Wesselink.

This level of discrepancy among very different methods, each of which seems well defined and persuasive on its own terms, is the crux of the current distance scale problem. Do the *Hipparcos* parallaxes in fact contain small and ill-understood errors of their own? Is it valid to apply Lutz-Kelker corrections – or more generally, other types of bias corrections – to single stars, or small numbers of them whose selection criteria are poorly determined and inhomogeneous? And how many of the subdwarfs are actually binaries?

The one subdwarf for which no luminosity bias correction is needed (or in dispute) is still Groombridge 1830 (HD 103095), by far the nearest one known. As an instructive numerical exercise, let us match this one star *alone* to the cluster M3 (NGC 5272), which has essentially the same metallicity and is also unreddened. Its *Hipparcos* measured parallax is $\pi = (109.2 \pm 0.8)$ mas, while the best ground-based compilation (from the Yale catalog; see van Altena et al. 1995) gives $\pi = (112.2 \pm 1.6)$ mas. The photometric indices for Gmb 1830, from several literature sources, are $V = 6.436 \pm 0.007$, $(B-V) = 0.75 \pm 0.005$, $(V-I) = 0.87 \pm 0.01$, giving $M_V = 6.633 \pm 0.016$ with no significant bias corrections. Its metallicity is [Fe/H] $= -1.36 \pm 0.04$ (from a compilation of several earlier studies) or -1.24 ± 0.07 from the data of Gratton et al. (1997).

This is nearly identical with [Fe/H] $= -1.34 \pm 0.02$ for M3 (Carretta & Gratton 1997). Gmb 1830 can safely be assumed to be unreddened, and the foreground reddening for M3 is usually taken as $E(B-V) = 0.00$ (Harris 1996a) and is in any case unlikely to be larger than 0.01. Thus the color adjustments to Gmb 1830 are essentially negligible as well. No other degrees of freedom are left, and we can match the star directly to the M3 main sequence at the same color to fix the cluster distance modulus. The result of this simple exercise is shown in Fig. 1.16. It yields $M_V(HB) = 0.59 \pm 0.05$, which is ~ 0.2 mag fainter than the level obtained by Reid (1997) or Gratton et al. (1997) from the entire sample of subdwarfs.

Clearly, it is undesirable to pin the entire globular cluster distance scale (and hence the age of the universe) on just one star, no matter how well determined. Nevertheless, this example illustrates the fundamental uncertainties in the procedure.

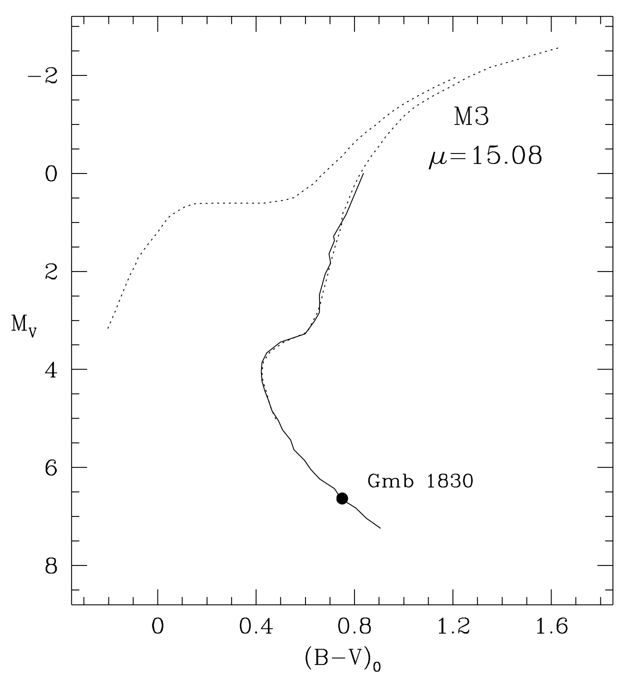

Fig. 1.16. Main sequence fit of the nearest subdwarf, Groombridge 1830, to the globular cluster M3. The cluster and the subdwarf have nearly identical metallicities and are unreddened. The solid line gives the deep main sequence and subgiant data for M3 from Stetson (1998), while the dotted line defining the brighter sections of the CMD is from Ferraro et al. (1997). The resulting distance modulus for M3 is $(m-M)_0 = 15.08 \pm 0.05$

1.2.7 A Synthesis of the Results for the Milky Way

The upper and lower extremes for the globular cluster distance scale as we now have them are well represented by the Baade-Wesselink field RR Lyrae calibration (Eq. 1.8) and the Gratton et al. *Hipparcos*-based subdwarf fits (Eq. 1.10). These are combined in Fig. 1.17 along with the results from the other selected methods listed above. Also notable is the fact that the slope of the relation is consistently near $\alpha \simeq 0.15$ (see also Carney in this volume). To set the zeropoint, I adopt a line passing through the obvious grouping of points near [Fe/H] ~ -1.4, and about halfway between the two extreme lines. This relation (solid line in Fig. 1.17) is

$$M_V(HB) = 0.15\,[\mathrm{Fe/H}] + 0.80\,. \tag{1.11}$$

Realistically, what uncertainty should we adopt when we apply this calibration to measure the distance to any particular object? Clearly the error is dominated not by the internal uncertainty of any one method, which is typically in the range $\pm 0.05 - 0.10$ mag. Instead, it is dominated by the external level of disagreement between the methods. How much weight one should put on any one method has often been a matter of personal judgement. As a compromise – perhaps a pessimistic one – I will use $\sigma(M_V) = \pm 0.15$ mag as an estimate of the true external uncertainty of the calibration at any metallicity.

A comprehensive evaluation of the distance scale, concentrating on the subdwarf parallax method but also including a long list of other methods, is given by Carretta et al. (1999). Their recommended HB calibration – tied in part to the distance to the LMC measured by both Population I and II standard candles – corresponds to $M_V(HB) = 0.13$ [Fe/H] $+0.76$, scarcely different from Eqn. 1.11 above. (NB: note again that $M_V(HB)$ is subtly different from both $M_V(ZAHB)$ and $M_V(RR)$: the ZAHB is roughly 0.1 mag fainter than the mean HB because of evolutionary corrections, and the mean level of the RR Lyraes is about 0.05 mag brighter than the ZAHB for the same reason. As noted previously, I use the mean HB level without adjustments.)

1.2.8 Comparisons in the LMC and M31

Extremely important external checks on the globular cluster distance scale can be made through the Cepheids and other Population I standard candles, once we go to Local Group galaxies where both types of indicators are found at common distances. By far the most important two "testing grounds" are the Large Magellanic Cloud and M31, where several methods can be strongly tested against one another.

For the LMC, RR Lyrae variables are found in substantial numbers both in its general halo field and in several old globular clusters. The field-halo variables have mean V magnitudes as listed in Table 1.4 below, from five

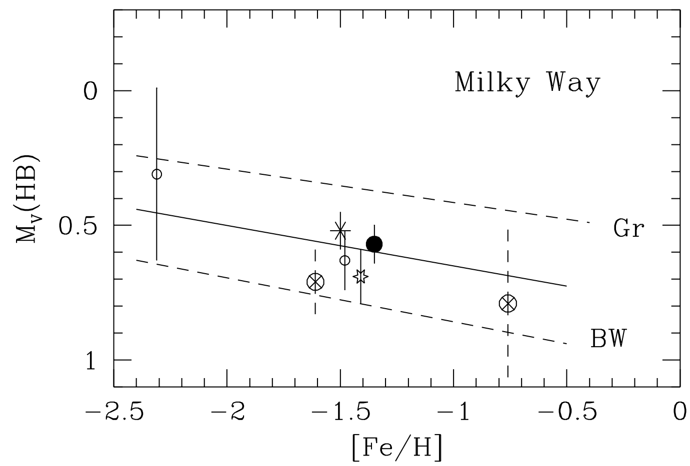

Fig. 1.17. Calibrations of the HB luminosity for Milky Way globular clusters. The upper dashed line (Gr) is the *Hipparcos* subdwarf calibration from Gratton et al. (1997), and the lower dashed line (BW) is the Baade-Wesselink calibration for field RR Lyraes from Carney et al. (1992), as listed in the text. Other symbols are as follows: *Solid dot:* Main sequence fit of Groombridge 1830 to M3. *Large asterisk:* Fit of white dwarf sequence in NGC 6752 to nearby field white dwarfs. *Small open circles:* Astrometric parallaxes, from Rees (1996) in two metallicity groups. *Large circled crosses:* Statistical parallax of field RR Lyrae stars, from Layden et al. (1996). *Open star:* Mean trigonometric parallax of field HB stars. Finally, the *solid line* is the adopted calibration, $M_V(HB) = 0.15$ [Fe/H] + 0.80

studies in which statistically significant numbers of variables have been measured. Using a foreground absorption for the LMC of $E(B-V) = 0.08 \pm 0.01$ and $A_V = 0.25 \pm 0.03$, I calculate a weighted mean dereddened magnitude $\langle V_0 \rangle = 18.95 \pm 0.05$ (the mean is driven strongly by the huge MACHO sample, though the other studies agree closely with it). The mean metallicity of the field variables appears to be near [Fe/H] $\simeq -1.7$ (see van den Bergh 1995, and the references in the table). Thus our adopted Milky Way calibration would give $M_V(RR) = 0.55 \pm 0.15$ (estimated external error) and hence a true distance modulus $(m-M)_0$(LMC) $= 18.40 \pm 0.15$.

Well determined mean magnitudes are also available for RR Lyrae stars in seven LMC globular clusters (Walker 1989; van den Bergh 1995). Using the same foreground reddening, we find an average dereddened RR Lyrae magnitude for these clusters of $\langle V_0 \rangle = 18.95 \pm 0.05$. Their mean metallicity in this case is [Fe/H] $\simeq -1.9$, thus from our Milky Way calibration we would predict $M_V(RR) = 0.52 \pm 0.15$ and hence $(m-M)_0$(LMC) $= 18.44 \pm 0.15$. The cluster and field RR Lyrae samples are in substantial agreement. Gratton

Table 1.4. Field RR Lyrae stars in the LMC

Location	$\langle V \rangle_{RR}$	Source
NGC 1783 field	19.25 ± 0.05	Graham 1977
NGC 2257 field	19.20 ± 0.05	Walker 1989
NGC 1466 field	19.34 :	Kinman et al. 1991
NGC 2210 field	19.22 ± 0.11	Reid & Freedman 1994
MACHO RRd's	19.18 ± 0.02	Alcock et al. 1997

(1998) and Carretta et al. (1999) suggest, however, that the *central bar* of the LMC could be at a different distance – perhaps as much as 0.1 mag further – than the average of the widely spread halo fields. Unfortunately, it is the LMC bar distance that we really want to have, so this contention introduces a further level of uncertainty into the discussion.

How do these RR Lyrae-based distance estimates compare with other independent standard candles, such as the LMC Cepheids or the SN1987A ring expansion? These methods themselves are not without controversy (for more extensive reviews, see, e.g., van den Bergh 1995; Fernley et al. 1998a; Gieren et al. 1998; or Feast 1998). Fundamental parallax distances to the Hyades and Pleiades can be used to establish main sequence fitting distances to Milky Way open clusters containing Cepheids, which then set the zeropoint of the Cepheid period-luminosity relation and hence the distance to the LMC. The Baade-Wesselink method can also be adapted to set distances to Cepheids in the nearby field. The SN1987A ring expansion parallax is an important new independent method, but here too there are disagreements in detail about modelling the ring geometry (cf. the references cited above). A brief summary of the most accurate methods, drawn from Fernley et al. (1998a) and Gieren et al. (1998), is given below in Table 1.5. Although the individual moduli for these methods (as well as others not listed here) range from ~ 18.2 up to 18.7, it seems to me that an adopted mean $(m-M)_0(\mathrm{LMC}) = 18.5 \pm 0.1$ is not unreasonable. For comparison, the comprehensive review of Carretta et al. (1999) recommends $(m-M)_0 = 18.54 \pm 0.04$.

Table 1.5. A summary of distance calibrations for the LMC

Method	$(m-M)_0$
Cepheids (Clusters, BW)	18.49 ± 0.09
Mira PL relation	18.54 ± 0.18
SN1987A ring (4 recent analyses)	18.51 ± 0.07
RR Lyraes (clusters, field)	18.44 ± 0.15

The step outward from the LMC to M31 can be taken either by comparing the mean magnitudes of the halo RR Lyrae variables in each galaxy, by the Cepheids in each, or by the RGB tip stars:

- *RR Lyraes*: In the M31 halo, the sample of RR Lyraes found by Pritchet & van den Bergh (1987) has $\langle V_0 \rangle$(M31) $= 25.04 \pm 0.10$ and thus $\Delta(m-M)_0$(M31-LMC) $= 6.09 \pm 0.11$, or $(m-M)_0 = 24.59 \pm 0.15$.
- *Cepheids*: Two studies employing optical photometry give $\Delta(m-M)_0 = 5.92 \pm 0.10$ (Freedman & Madore 1990) or 6.07 ± 0.05 (Gould 1994) from different prescriptions for matching the P-L diagrams in the two galaxies. The recent study of Webb (1998), from JHK near-infrared photometry which is less affected by reddening and metallicity differences, gives $\Delta(m-M)_0 = 5.92 \pm 0.02$ and thus $(m-M)_0 = 24.42 \pm 0.11$.
- *Red Giant Branch Tip*: A precise method which is more or less independent of both the Cepheids and RR Lyraes is the luminosity of the old red giant branch tip (TRGB) of the halo stars (essentially, the luminosity of the core helium flash point), which for metal-poor populations has a nearly constant luminosity $M_I = -4.1 \pm 0.1$ (Lee et al. 1993a; Harris et al. 1998b). For a wide sample of the M31 halo field giants, Couture et al. (1995) find an intrinsic distance modulus $(m-M)_0 = 24.5 \pm 0.2$.
- *Other methods*: Other useful techniques include the luminosities of disk carbon stars (Brewer et al. 1995), surface brightness fluctuations of globular clusters (Ajhar et al. 1996), and luminosities of the old red clump stars (Stanek & Garnavich 1998). These give results in exactly the same range $(m-M)_0 \sim 24.4 - 24.6$. Holland (1998) has used theoretical isochrone fits to the red giant branches of 14 halo globular clusters in M31 to obtain $(m-M)_0 = 24.47 \pm 0.07$.

Combining all of these estimates, I will adopt $(m-M)_0$(M31) $= 24.50 \pm 0.14$. Putting back in the M31 foreground reddening $E(B-V) = 0.09 \pm 0.03$ (van den Bergh 1995) then gives an apparent distance modulus $(m-M)_V$(M31)$=$ 24.80 ± 0.15.

To bring a last bit of closure to our discussion, we can finally test our Milky Way globular cluster distance scale against the mean HB levels that are directly observed in the globular clusters of M31. Fusi Pecci et al. (1996) have carried out homogeneous reductions for *HST* images of seven M31 clusters, with the results as shown in Fig. 1.18. The unweighted least-squares line defined by these seven points[5] is

$$V_{HB}(\text{observed}) = (0.096 \pm 0.078)\,[\text{Fe/H}] \; + \; (25.56 \pm 0.09)\,. \tag{1.12}$$

[5] This is *not* the same line derived by Fusi Pecci et al.; both their slope and zeropoint are slightly different. The reason for the difference is that they adjust the raw V_{HB} values to the somewhat fainter unevolved ZAHB position. Here, I use the directly observed V_{HB} without adjustment.

For comparison, if we take our fiducial Milky Way relation and transport it outward by our best-estimate distance modulus $(m-M)_V = 24.8$, we obtain

$$V_{HB}(\text{predicted}) = M_V(HB) + 24.80 = 0.15\,[\text{Fe/H}] + (25.60 \pm 0.15)\,. \tag{1.13}$$

These two relations are remarkably consistent with one another, and give additional confidence that our fundamental distance scale is not likely to be wrong by more than the tolerance that we have claimed.

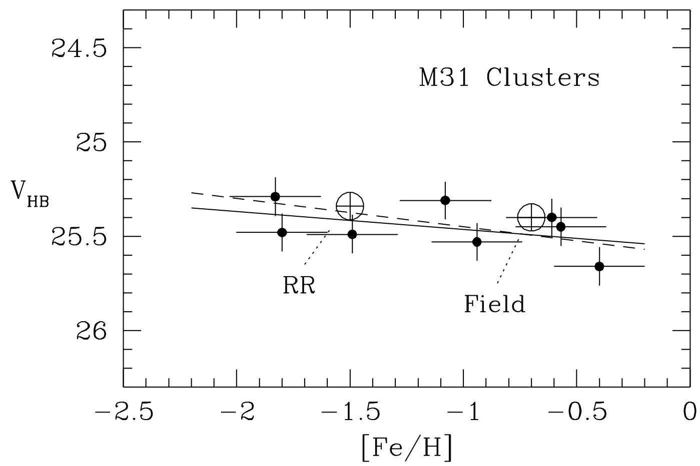

Fig. 1.18. Horizontal branch levels for globular clusters in M31, plotted against cluster metallicity. *Solid dots:* V_{HB} values for seven clusters from Fusi Pecci et al. (1996). *Large circled crosses:* Mean magnitudes for the halo field RR Lyraes (van den Bergh 1995) and the red HB stars near cluster G1 (Rich et al. 1996). The *solid line* shows the mean relation for the seven clusters as defined in the text, while the *dashed line* is our fiducial Milky Way relation added to the distance modulus $(m-M)_V(\text{M31}) = 24.80$

1.3 THE MILKY WAY SYSTEM: KINEMATICS

A hypothesis or theory is clear, decisive, and positive, but it is believed by no one but the man who created it. Experimental findings, on the other hand, are messy, inexact things which are believed by everyone except the man who did that work.

Harlow Shapley

Much information about the origin and history of the Milky Way GCS is contained in the cluster space motions or *kinematics*. Armed with this kind of information along with the spatial distributions and cluster metallicities (§1), we can make considerably more progress in isolating recognizable subsystems within the GCS, and in comparing the clusters with other types of halo stellar populations.

The seminal work in kinematics of the GCS is to be found in the pioneering study of Mayall (1946), a paper which is just as important in the history of the subject as the work by Shapley (1918) on the cluster space distribution.[6] Other landmarks that progressively shaped our prevailing view of the GCS kinematics are to be found in the subsequent work of Kinman (1959a,b), Frenk & White (1980), and Zinn (1985).

1.3.1 Coordinate Systems and Transformations

The basic question in GCS kinematics is to determine the relative amounts of *ordered* motion of the clusters (net systemic rotation around the Galactic center) and *random* internal motion. The ratio of these quantities must depend on their time and place of origin, and thus on such measurables as cluster age, spatial location, or metallicity.

The first attempts at kinematical solutions (Mayall, Kinman, and others) used the simplest traditional formalism in which the Solar motion $\boldsymbol{U}$ was calculated relative to the clusters or various subsets of them. Formally, if v_r equals the radial velocity of the cluster relative to the Solar Local Standard of Rest, then

$$v_r = U \cos\lambda \tag{1.14}$$

where λ is the line-of-sight angle to the cluster (defined in Fig. 1.19 below). A graph of v_r against $(\cos \lambda)$ should yield a straight-line solution through the origin with slope U. For true halo objects with little or no systemic rotation, U must therefore be approximately equal to V_0(LSR), the rotation speed of the Solar Local Standard of Rest around the Galactic center; more strictly, U relative to the GCS must represent a lower limit to V_0 except in extreme scenarios where the halo, or part of it, is in retrograde rotation. Mayall, in the very first attempt to do this, derived $U \simeq 200$ km s^{-1}, a value scarcely different from the modern solutions of $\simeq 180$ km s^{-1}. The simple Solar motion approach, however, gives little information about the net motions of the many clusters for which the line of sight from Sun to cluster is roughly at right angles to the $\boldsymbol{V_0}$ vector. Instead, we will move directly on to the modern formalism, as laid out (e.g.) in Frenk & White (1980) and in many studies since.

[6] Mayall's paper is essential reading for any serious student of the subject. Now half a century old, it stands today as a remarkable testament to the author's accomplishment of a major single piece of work in the face of several persistent obstacles. It also typifies a brutally honest writing style that is now rather out of fashion.

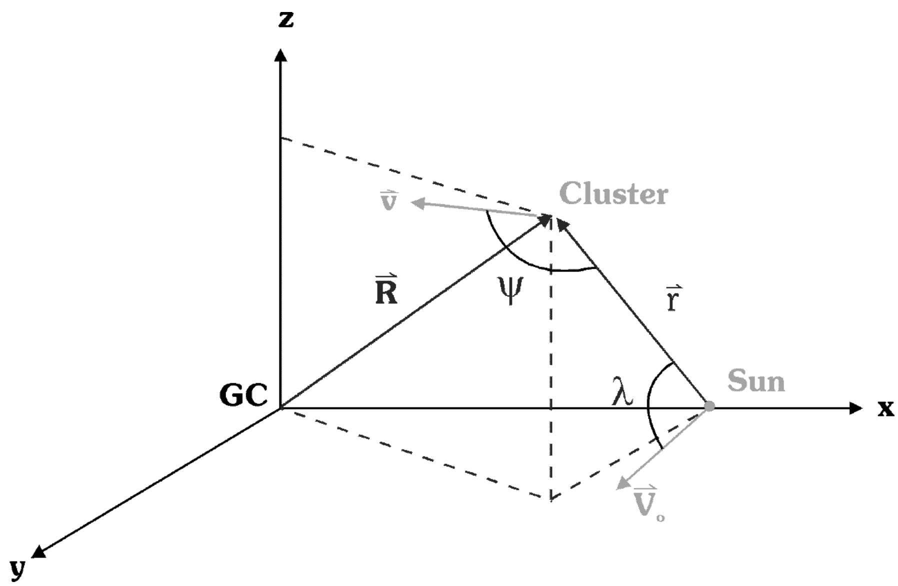

Fig. 1.19. Geometry for the rotational motions of Sun S and cluster C around the Galactic center GC

Referring to Fig. 1.19, let us consider a cluster which has a Galactocentric distance R and a rotation speed $V(R)$ around the Galactic center. Its radial velocity relative to the Solar LSR is then

$$v_r = V \cos\psi \; - \; V_0 \cos\lambda \tag{1.15}$$

where λ is the angle between the Solar motion vector $\boldsymbol{V_0}$ and the line of sight $\boldsymbol{r}$ to the cluster; and ψ is the angle between $\boldsymbol{r}$ and the rotation vector $\boldsymbol{V}$ of the cluster. We have

$$\boldsymbol{V_0} \cdot \boldsymbol{r} \; = \; V_0 \, r \cos\lambda \tag{1.16}$$

which gives, after evaluating the dot product,

$$\cos\lambda \; = \; \cos b \; \cdot \; \sin\ell \, . \tag{1.17}$$

Similarly, cos ψ can be evaluated from the dot product

$$\boldsymbol{V} \cdot \boldsymbol{r} \; = \; V \, r \cos\psi \tag{1.18}$$

which eventually gives

$$\cos\psi = \frac{R_0 \cos b \sin\ell}{((r \cos b \sin\ell)^2 \; + \; (R_0 - r \cos b \cos\ell)^2)^{1/2}} \, . \tag{1.19}$$

The equation of condition for V is then

$$V\cos\psi = v_r + V_0\cos\lambda \tag{1.20}$$

where v_r is the directly measured radial velocity of the cluster (relative to the LSR!); and (λ, ψ) are known from the distance and direction of the cluster. We explicitly *assume* $V_0 = 220$ km s^{-1} for the Solar rotation. The quantity on the right-hand side of the equation is the radial velocity of the cluster relative to a stationary point at the Sun (i.e., in the rest frame of the Galactic center). Thus when it is plotted against cos ψ, we obtain a straight-line relation with slope V (the net rotation speed of the group of clusters) and intercept zero.

Frenk & White (1980) demonstrate that an unbiased solution for V is obtained by adding the weighting factor cos ψ,

$$V = \frac{\langle\cos\psi(v_r + V_0\cos\lambda)\rangle}{\langle\cos^2\psi\rangle} \pm \frac{\sigma_{los}}{(\Sigma\cos^2\psi)^{1/2}} \tag{1.21}$$

where σ_{los}, the "line of sight" velocity dispersion, is the r.m.s. dispersion of the data points about the mean line.

Which of the parameters in the above equation can potentially generate significant errors in the solution? As we will see below, typically $\sigma_{los} \sim 100$ km s^{-1}, whereas the measurement uncertainties in the radial velocities of the clusters are $\epsilon(v_r) \lesssim 5$ km s^{-1} (cf. Harris 1996a). The radial velocity measurements themselves thus do not contribute anything important to uncertainties in V or σ_{los}. In addition, λ depends only on the angular location of the cluster on the sky and is therefore virtually error-free. The last input parameter is the angle ψ: *Uncertainties in the estimated distances* r can affect ψ severely – and asymmetrically – as is evident from inspection of Fig. 1.19, and these can then be translated into biases in V and σ_{los}. This point is also stressed by Armandroff (1989). In turn, the uncertainty $\epsilon(m-M)_V$ in distance modulus is most strongly correlated with cluster reddening (larger reddening increases both the absolute uncertainty in the absorption correction A_V, and the amount of *differential reddening*, which makes the identification of the CMD sequences less precise). A rough empirical relation

$$\epsilon(m-M)_V \simeq 0.1 + 0.4\,E(B-V) \tag{1.22}$$

represents the overall effect reasonably well.

Clearly, the clusters near the Galactic center and Galactic plane – which preferentially include the most metal-rich clusters – will be the most severely damaged by this effect. Since the dependence of ψ on r is highly nonlinear, the error bars on cos ψ can be large and asymmetric for such clusters. An example is shown in the kinematics diagram of Fig. 1.20. By contrast, the same diagram for any subset of the high-halo clusters is, point by point, far more reliable and thus considerably more confidence can be placed in the (V, σ_{los}) solution for such subgroups. Fortunately, however, the high numerical weights given to the clusters at large (cos ψ), which are exactly the objects that have the lowest reddenings and the most reliable distance estimates, make the solution for rotation speed V more robust than might at first be expected.

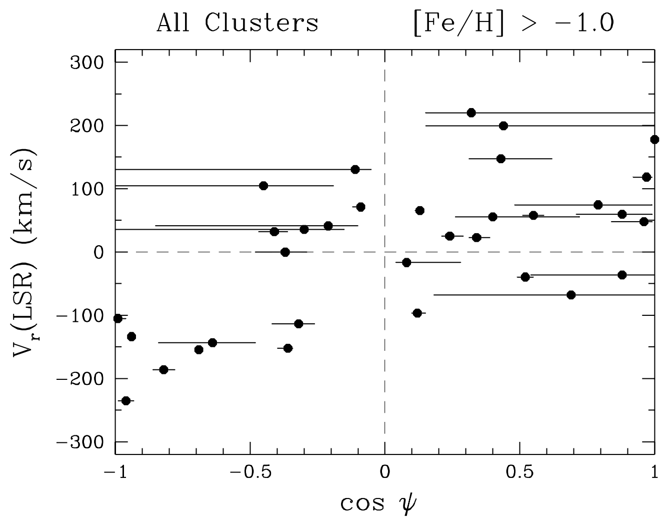

Fig. 1.20. Kinematics diagram for the metal-rich clusters in the Milky Way. Here $v_r(\mathrm{LSR}) = v_r + V_0 \cos \lambda$ is the radial velocity of the cluster relative to a stationary point at the Solar LSR. The horizontal error bars on each point show how each cluster can shift in the diagram due solely to uncertainties in its estimated distance, as given by Eq. 1.22

1.3.2 The Metal-Rich Clusters: Bulge-Like and Disk-Like Features

Somewhat contrary to historical tradition, let us first investigate the kinematics of the MRC clusters.

In Fig. 1.21, we see the kinematics diagrams for the inner ($R_{gc} < 4$ kpc) and outer (4 kpc $< R_{gc} <$ 9 kpc) MRC clusters. The best-fit numerical solutions, listed in Table 1.6, show healthy rotation signals and moderately low rms dispersions for both, although $V(\mathrm{rot})$ is clearly higher (and σ_{los} lower) for the outer sample. The inner subgroup is what we discussed in Section 1 as Minniti's (1995) bulge-like population. The velocities for these clusters are replotted against Galactic longitude in Fig. 1.22, following Minniti (1995) and Zinn (1996), in which it can be seen that they match well with the net rotation speed of the RGB stars in the bulge. When we add this evidence (not conclusive by itself!) to the space distribution discussed in Section 1, it seems likely that the inner MRC clusters are plausibly interpreted as a flattened bulge population with a rotation speed near $V \sim 90$ km s^{-1}.

The outer subgroup ($4 - 9$ kpc; second panel of Fig. 1.21) more nearly resembles what Zinn (1985) and Armandroff (1989) first suggested to be a "thick disk" population. The issue is discussed at length in other recent papers by Armandroff (1993), Norris (1993), and Zinn (1996). If this identification is correct, it would be highly suggestive that there is a genuine disk subsystem

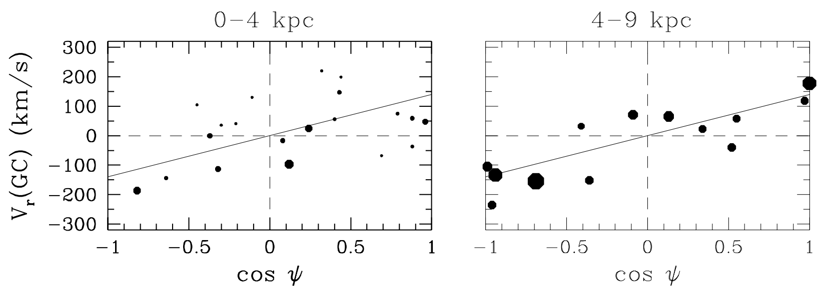

Fig. 1.21. Kinematics diagrams for the metal-rich Milky Way clusters. *First panel:* Normalized radial velocity against position angle for the innermost MRC clusters (R_{gc} < 4 kpc). The relative statistical weights for each point are indicated by symbol size (see text). *Second panel:* Outer MRC clusters (4 kpc < R_{gc} < 9 kpc). Note the larger weights on these points because of their lower reddenings and more accurate distance measurements. The same solid line (V = 140 km s^{-1} rotation speed) is plotted in both panels, although the formal solution for the innermost clusters is $V \sim 86$ km s^{-1}; see the text

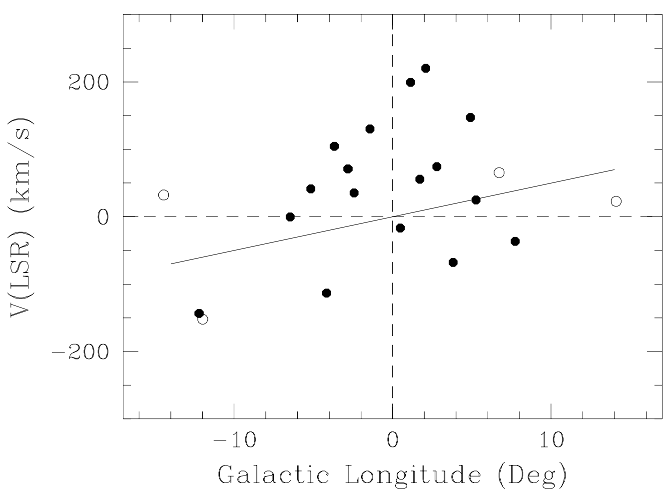

Fig. 1.22. Normalized radial velocity against Galactic longitude, for the MRC clusters within $\sim 15^o$ of the Galactic center. Filled circles are clusters within 4 kpc of the center; open circles are ones with R_{gc} > 4 kpc. The solid line represents the rotation curve of the Galactic bulge from red giant stars (Minniti 1995; Zinn 1996)

within the Milky Way GCS which formed along with the thick-disk stars; if so, it should then be possible to set the formation epoch of the thick disk quite accurately by the chronology of these clusters.

Although this interpretation of the data is well known, it is not quite ironclad. The well determined rotation speed of the outer MRC clusters, $V = 147 \pm 27$ km s^{-1}, is noticeably less than normally quoted values for the thick-disk stars, which are near $V \simeq 180$ km s^{-1} (cf. Armandroff 1989; Norris 1993; Majewski 1993). It is tempting to imagine that the outer MRC clusters may be the remnants of a "pre-thick-disk" epoch of star formation, during which their parent Searle-Zinn gaseous fragments had not fully settled into a disklike configuration, still preserving significant random motions. The leftover gas from this period would have continued to collapse further into the thick disk and (later) the old thin disk, with progressively larger rotation speeds.

Table 1.6. Mean rotation velocities of subsets of clusters

Sample	Subgroup	n	V (km/s)	σ_{los} (km/s)
MRC	All [Fe/H] > -1	33	118 ± 26	89 ± 11
MRC	$R_{gc} = 0 - 4$ kpc	20	86 ± 40	99 ± 15
MRC	$R_{gc} = 4 - 9$ kpc	13	147 ± 27	66 ± 12
MPC	All [Fe/H] < -1	89	30 ± 25	121 ± 9
MPC	$R_{gc} = 0 - 4$ kpc	28	56 ± 37	122 ± 16
MPC	$R_{gc} = 4 - 8$ kpc	19	12 ± 31	79 ± 12
MPC	$R_{gc} = 8 - 12$ kpc	12	26 ± 63	148 ± 29
MPC	$R_{gc} = 12 - 20$ kpc	14	-97 ± 110	132 ± 24
MPC	$-2.30 <$ [Fe/H] < -1.85	17	139 ± 57	114 ± 19
MPC	$-1.85 <$ [Fe/H] < -1.65	19	41 ± 55	142 ± 22
MPC	$-1.65 <$ [Fe/H] < -1.50	21	-35 ± 59	134 ± 20
MPC	$-1.50 <$ [Fe/H] < -1.32	17	-12 ± 56	106 ± 17
MPC	$-1.32 <$ [Fe/H] < -1.00	17	31 ± 32	80 ± 13
MPC	All [Fe/H] < -1.70	30	80 ± 43	130 ± 16
MPC	BHB, $R_{gc} > 8$ kpc	20	55 ± 58	115 ± 17
MPC	RHB, $R_{gc} > 8$ kpc	18	-39 ± 83	158 ± 26
MPC	RHB excl. N3201	17	32 ± 88	149 ± 24

At the same time, it is plainly true that some individual clusters have disk-like orbital motions (e.g., Cudworth 1985; Rees & Cudworth 1991; Dinescu et al. 1999). Burkert & Smith (1997) have gone further to suggest that the outer MRC clusters form a disklike subsystem, while the inner low-luminosity ones form an elongated bar-like structure (see also Côté 1999). However, a serious concern is that the distance estimates for the inner low-luminosity bulge clusters are likely to be more badly affected by extreme reddening and field contamination than for the luminous clusters in the same region of the bulge. The apparent elongation of the set of low-luminosity clusters along the X axis may therefore be an artifact. Better photometry and cleaner CMDs for these objects, possibly from near-infrared observations, are needed to clear up the problem.

Why should we try so hard to relate the globular clusters to the halo field stars? There is a nearly irresistible temptation to force some given subset of the globular clusters to correspond *exactly* with some population of field stars that look similar according to their kinematics, metallicity distribution, and space distribution. But the more we find out about these subsystems in detail, the harder it is to make such precise correpondences. As will be discussed later (Section 8), the formation of massive star clusters must be a relatively rare and inefficient process within their progenitor gas clouds. If the globular clusters formed first out of the densest gas clumps, the remaining gas (which in fact would be the majority of the protostellar material) would have had plenty of opportunity to collide with other gaseous fragments, dissipate energy, and take up new configurations before forming stars. By then, it would have lost its "memory" of the earlier epoch when the globular clusters formed, and would essentially behave as a different stellar population (Harris 1998). In short, there seems to be no compelling reason to believe that subgroups of the GCS should be cleanly identified with any particular field-star population. This concern will surface again in the next section.

1.3.3 The Metal-Poor Clusters

We will now turn to the MPC clusters, which form the majority of the Milky Way globular cluster system. One minor correction we need to make before proceeding is to note that the four clusters believed to belong to the Sagittarius dwarf (NGC 6715, Arp 2, Ter 7, Ter 8; see Da Costa & Armandroff 1995) all have similar space motions and locations: we will keep only NGC 6715 as the "elected representative" for Sagittarius and discard the other three. Other correlated moving groups involving similarly small numbers of clusters have been proposed to exist (Lynden-Bell & Lynden-Bell 1995; Fusi Pecci et al. 1995), but these are much less certain than Sagittarius, and for the present we will treat the remaining clusters as if they are all uncorrelated.

Plotting all the MPC clusters at once in the kinematics diagram (first panel of Fig. 1.23), we see that as a whole it is totally dominated by random motion with no significant mean rotation. As before, the symbol size denotes the relative uncertainty in cos ψ; since most of the objects here have low reddenings and well determined distances, they have much more accurately fixed locations in the diagram.

However, throwing them all into the same bin is guaranteed to obscure the existence of any distinct subsystems. The other panels of Fig. 1.23 show the sample broken into five bins of Galactocentric distance, with rather arbitrarily chosen boundaries. Again, the individual solution parameters are listed in Table 1.6. A small positive rotation signal is present in the innermost 4-kpc zone, whereas no significant rotation appears in any of the other bins. As is evident from the figure, for clusters more distant than $R_{gc} \gtrsim 15$ kpc the range of $\cos\psi$ becomes so small that no valid solution for V can be performed. In the $4-8$ kpc bin, the dispersion is distinctly lower than in any of the other bins; the meaning of this anomaly is unclear (particularly since it is accompanied by zero rotation), and the possibility that it is simply a statistical fluctuation cannot be ruled out given the small numbers of points.

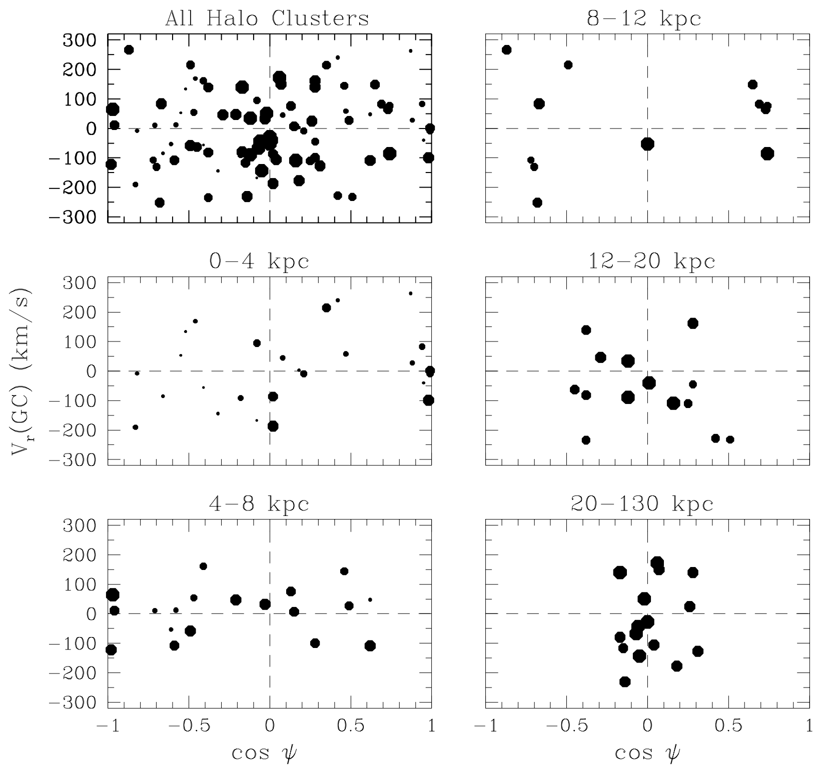

Fig. 1.23. Kinematics diagrams for the metal-poor clusters, divided into Galactocentric distance intervals. Larger points indicate clusters with lower reddenings and thus better determined distances and ψ values

Grouping the clusters by metallicity is also instructive. In most Galaxy formation models, we might anticipate that this version would be closer to a chronological sequence where the higher-metallicity objects formed a bit later in the enrichment history of their parent gas clouds. A sample of this breakdown is shown in Fig. 1.24, where now we exclude the six most remote clusters ($R_{gc} > 50$ kpc) that have no effect on the solution. *No significant net rotation is found* for any metallicity subgroup except for the lowest-metallicity bin. For the intermediate-metallicity groups, notice (Table 1.6) the slight (but not statistically significant!) dip into net retrograde rotation. The middle bin in particular is influenced strongly by the single object NGC 3201 (point at uppermost left). This particular cluster has a uniquely strong influence on the kinematical solution because of its location near the Solar antapex and large positive radial velocity; it carries the highest statistical weight of any cluster in the entire sample. More will be said about this interesting and somewhat deceptive subgroup below.

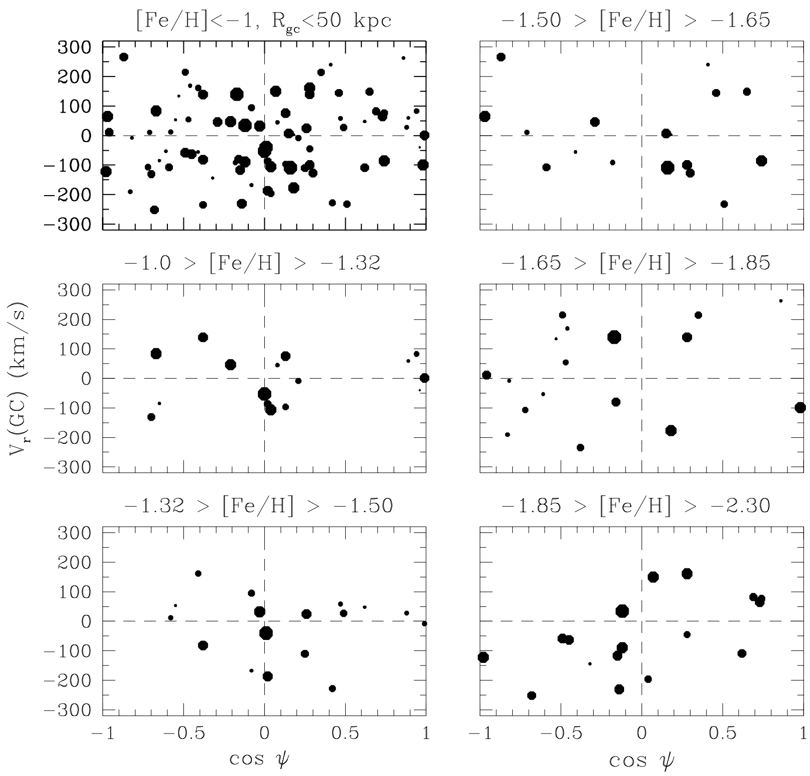

Fig. 1.24. Kinematics diagrams for the metal-poor clusters, divided by [Fe/H] intervals. Larger points indicate clusters with lower reddenings and thus better determined distances and ψ values

A potentially more important trend, which does not depend on a single object, shows up in the very lowest metallicity bin. The metal-poorest clusters exhibit a strong and significant net rotation. To find out where this signal is coming from, in Fig. 1.25 we combine all clusters more metal-poor than

[Fe/H] = −1.7 and relabel them by distance as well. As noted above, we find that the objects with $R_{gc} \gtrsim 15$ kpc contribute little to the solution for V (but do affect the dispersion); it is the mid- to inner-halo objects which drive the rotation solution, even though some have low weight because of uncertain distances. The result from Fig. 1.25 is $V = (80 \pm 43)$ km s^{-1} – a surprisingly large positive value, considering that we would expect this set of clusters to be the oldest ones in the Galaxy, and that no other subgroup of low-metallicity clusters displays any significant net rotation. The analysis of three-dimensional space motions by Dinescu et al. (1999) for a selection of these clusters yields a similar result. They obtain $V = (114 \pm 24)$ km s^{-1} for the metal-poor clusters with $R < 8$ kpc, formally in agreement with the solution from the radial velocities alone.

It seems necessary to conclude that the inner halo (0 – 4 kpc) has a rotation speed of $\sim 80 - 100$ km s^{-1} *regardless of metallicity:* the MPC and MRC clusters move alike. Are we seeing here the traces of an ELS-style formation epoch of collapse and spin-up in the inner halo, which (as SZ first claimed) would have been less important further out in the halo?

Still another way to plot this trend is shown in Fig. 1.26. Here, we start with the list of all 94 clusters with [Fe/H] < -0.95 and known velocities, sort them in order of metallicity, and solve for rotation V using the first 20 clusters in the list. We then shift the bin downward by one object (dropping the first one in the list and adding the 21st) and redo the solution. We shift the bin down again, repeating the process until we reach the end of the list. Clearly any one point is not at all independent of the next one, but this moving-bin approach is an effective way to display any global trends with changing metallicity. What we see plainly is the clear net rotation of the lowest-metallicity subpopulation, which smoothly dies away to near-zero rotation for [Fe/H] $\gtrsim -1.7$. (NB: The apparently sudden jump into retrograde rotation at [Fe/H] $\simeq -1.6$ is, once again, due to NGC 3201, which enters the bin in that range. If this cluster is excluded, the net rotation stays strictly near zero across the entire range.)

A minor additional point (shown in the lower panel of Figure 1.26) is that the mean galactocentric radius decreases slightly as the metallicity of the bin shifts from the most metal-poor objects to the less metal-poor end. This trend is simply the result of the fact that there is a small metallicity gradient in the MPC system (Section 1), with the most metal-poor objects located more frequently at larger radii.

1.3.4 Retrograde Motion: Fragments and Sidetracks

Though the ELS-style formation picture may still hold some validity for the Galaxy's inner halo ($R_{gc} \lesssim R_0$), very different ideas began to emerge for the outer halo especially in the literature of the past decade. Numerous pieces of evidence, as well as theoretical ideas, arose to suggest that much of the halo might have been *accreted* in the form of already-formed satellite fragments,

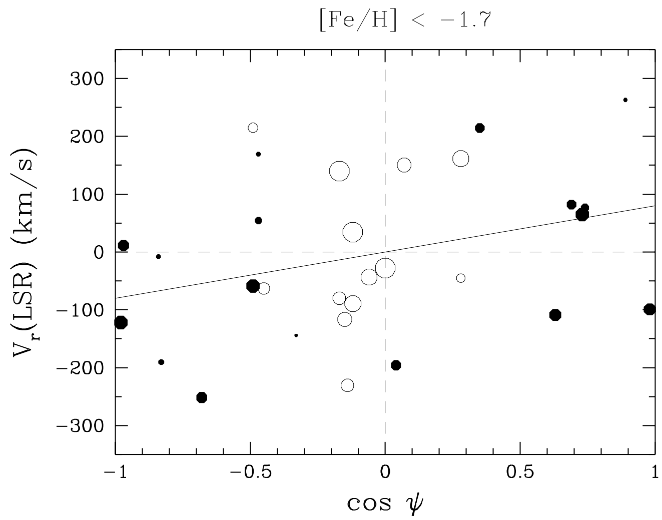

Fig. 1.25. Kinematics diagram for the extreme low metallicity clusters ([Fe/H] < -1.7). The solid line indicates the formal solution $V = 80$ km s^{-1} for this group of objects. Filled symbols are ones within 15 kpc of the Galactic center, open symbols are ones outside 15 kpc

each one of which would now be stretched out around the halo in a thin tidal streamer (see, for example, Majewski et al. 1996; Johnston 1998; Grillmair 1998, for review discussions with extensive references).

The particular relevance of these ideas to the globular clusters began with a comment by Rodgers & Paltoglou (1984) that the clusters in the metallicity range $-1.4 >$ [Fe/H] > -1.7 not only had a small anomalous retrograde rotation, but that most of them also had similar horizontal-branch morphologies. These objects included clusters like M3, NGC 3201, NGC 7006, and several others with HBs that are well populated across the RR Lyrae instability strip. By contrast, intermediate-metallicity clusters like M13 (with extreme blue HBs) did not show this collective retrograde rotation. Rodgers & Paltoglou speculated that the "anomalous" group might have had a common origin in a small satellite galaxy that was absorbed by the Milky Way on a retrograde orbit.[7] They suggested that by contrast, the M13-type clusters with prograde or near-zero rotation were the "normal" ones belonging to the

[7] It should be noted that *individual* clusters with retrograde orbits are certainly not unusual: since the halo velocity dispersion is high and it is basically a pressure-supported system (high random motions and low overall rotation speed), there will be a large mix of both prograde and retrograde orbits to be found. The issue here is that it is hard to see how a *collective* retrograde motion of an entire identifiable group of clusters could have arisen in any other way than accretion after the main *in situ* star formation phase of the halo.

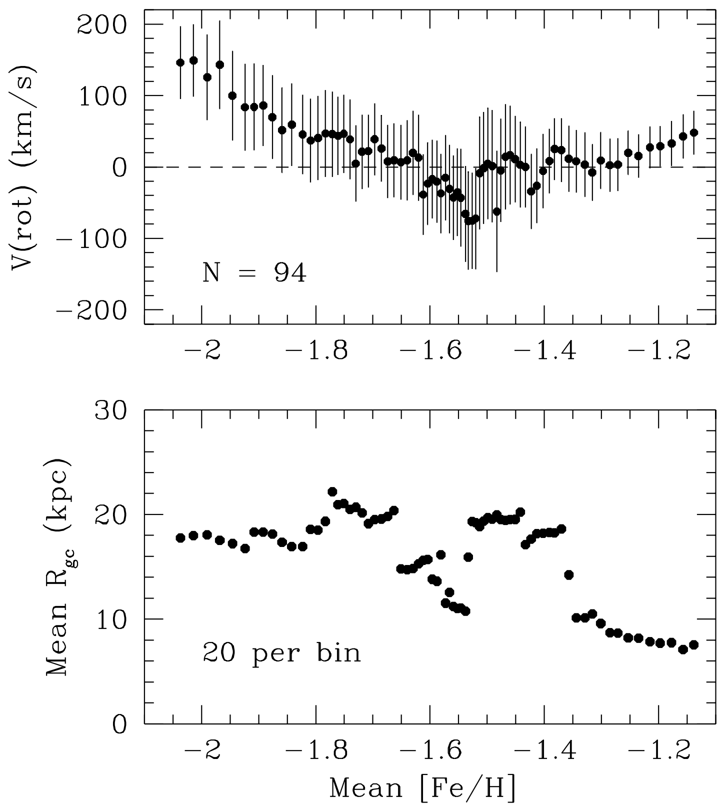

Fig. 1.26. Rotation speed plotted against mean [Fe/H], for the "moving bin" calculation described in the text. Over the interval $-2.3 \lesssim$ [Fe/H] $\lesssim -1.7$, note the smooth decline in V. Each bin contains 20 clusters

Milky Way halo from the beginning. The broader idea extending beyond this particular subset of clusters was that many individual ancestral satellites of the Galaxy might exist, and might still be identified today: to quote Rodgers & Paltoglou, "To be identified now as a component of the galactic outer halo, a parental galaxy must have produced a significant number of clusters in which a small range of metallicity is dominant and must have sufficiently distinct kinematics".

This idea was pursued later in an influential paper by Zinn (1993a) and again by Da Costa & Armandroff (1995). To understand it, we need to refer back to the HB morphology classification diagram of Fig. 1.9. In this diagram, the "normal" relation between HB type and metallicity is defined by the objects within $R_{gc} \lesssim 8$ kpc. The M3-type clusters further out in the halo and with generally redder HBs fall to the left of this normal line. The Rodgers & Paltoglou sample is drawn from the metallicity range -1.4 to -1.7, and indeed it can be seen that most of the clusters in that narrow horizontal cut across Fig. 1.9 belong to the red-HB group. Zinn, using the *assumption* that HB morphology is driven primarily by age for a given [Fe/H], called the normal ([Fe/H] < -0.8, blue-HB) clusters the Old Halo and the redder-HB ones the Younger Halo (though the latter group is not intended to be thought of as "young" in an absolute sense). If interpreted this way through typical HB models (e.g. Lee et al. 1994) – that is, if age is the dominant second parameter in Fig. 1.9 – then the Younger Halo clusters would need to be anywhere from ~ 2 to 5 Gyr younger than the Old Halo.

Zinn compared the kinematics of these two groups, finding $V = -64 \pm 74$ km s^{-1} for the Younger Halo and $V = 70 \pm 22$ km s^{-1} for the Old Halo, as well as a noticeably lower dispersion σ_{los} for the Old Halo. Developing the SZ formation picture further from these results, Zinn concluded "It seems likely ... that some of the outlying [protogalactic, gaseous] fragments escaped destruction, remained in orbit about the collapsed Galaxy, and evolved into satellite dwarf galaxies ... it is proposed that such satellite systems were the sites of the formation of the Younger Halo clusters".

Almost simultaneously, van den Bergh (1993a,b) used a different type of graphical analysis of kinematics to isolate rather similar subgroups, one of which (corresponding roughly to the Zinn Younger Halo) he postulates to have retrograde-type orbits. Van den Bergh went even further along the same line to envisage a single large ancestral fragment for these: "... the hypothetical ancestral galaxy that formed ... clusters with M3-like color-magnitude diagrams merged with the main body of the protoGalaxy on a plunging retrograde orbit".

It appears to me that these interpretations are quite risky, and that we need to take a fresh look at the actual data upon which they are built. There are at least two serious problems:

- The interpretation of HB morphology as a fair indication of cluster age (Lee et al. 1994; Chaboyer et al. 1996) is not proven; in fact, more recent evidence based directly on deep main-sequence photometry of clusters in each of the two groups suggests just the opposite in at least some cases. The clusters M3 and M13, with similar chemical compositions and different HB morphologies, form a classic "second parameter" pair. Precise differential main sequence fitting (Catelan & de Freitas Pacheco 1995; Johnson & Bolte 1998; Grundahl et al. 1998) indicates that these clusters have the same age to within the ~ 1 Gyr level which is the current precision of the technique. The still more extreme second-parameter trio NGC 288/362/1851 (e.g., Stetson et al. 1996; Sarajedini et al. 1997) may exhibit an age range of ~ 2 Gyr. Other recent studies of red-HB clusters in the outermost Milky Way halo (Stetson et al. 1999) and in the LMC and Fornax dwarf (e.g., Johnson et al. 1999; Olsen et al. 1998; Buonanno

et al. 1998) with moderately low metallicities show age differences relative to M3 and M92 that are modest, at most 2 Gyrs and often indistinguishable from zero. In short, the hypothesis that cluster age is the dominant second parameter may indeed work for some clusters but does not seem to be consistent with many others. Additional combinations of factors involving different ratios of the heavy elements, mixing, rotation, helium abundance, or mass loss will still need to be pursued much more carefully (cf. the references cited above).
We should not continue to use the terms "old" and "young" for these groups of clusters; in what follows, I will refer to them instead as the blue-HB and red-HB groups.
- The "retrograde rotation" of the red-HB group (Table 1.6) is not statistically significant. We also need to ask how it arises in the first place. The formally negative rotation of the red-HB group is driven *very strongly* by the single object NGC 3201, which (as we saw above) has a uniquely powerful influence on the kinematical solutions for any group it is put into. Taking NGC 3201 out of the sample (see Table 1.6) turns out to change $V(\mathrm{rot})$ by a full $+70$ km s^{-1}, changing the retrograde signature to a prograde one.[8] Neither the prograde or retrograde value is, however, significantly different from zero.

Some of these points are demonstrated further in Fig. 1.27. Here, we show the kinematics diagrams for the two groups of clusters. To make the groups as strictly comparable as possible, we draw each one *strictly from the same zone of the halo*, 8 kpc $< R_{gc} <$ 40 kpc, and ignore the blue-HB clusters in the inner halo. We see from the figure that neither sample has a significant rotation, either prograde or retrograde, whether or not we choose to remove NGC 3201 (though it is evident from the graphs just how influential that one cluster is).

True retrograde orbits are notoriously hard to deduce from radial velocities alone (NGC 3201 is one of the rare exceptions). This type of analysis would benefit greatly from reliable knowledge of *absolute proper motions* of these clusters, from which we can deduce their full three-dimensional space motions. Proper motions (μ_δ, μ_α) now exist in first-order form for almost 40 clusters, from several recent studies notably including Cudworth & Hansen (1993); Odenkirchen et al. (1997); and Dinescu et al. (1999); see Dinescu et al. for a synthesis of all the current results with extensive references. We can employ these to make useful classifications of orbital types (clearly prograde, clearly retrograde, or plunging) and the general range of orbital eccentricities. In the new orbital data summarized by Dinescu et al. (1999), we find 10 BHB

[8] We can, of course, treat the BHB group similarly by removing the single most extreme point (in this case, NGC 6101) and redoing the solution. $V(\mathrm{rot})$ changes from (55 ± 58) to (88 ± 54) km s^{-1}, a statistically insignificant difference. This test verifies again that NGC 3201 has a uniquely strong influence on whatever set of objects it is included with.

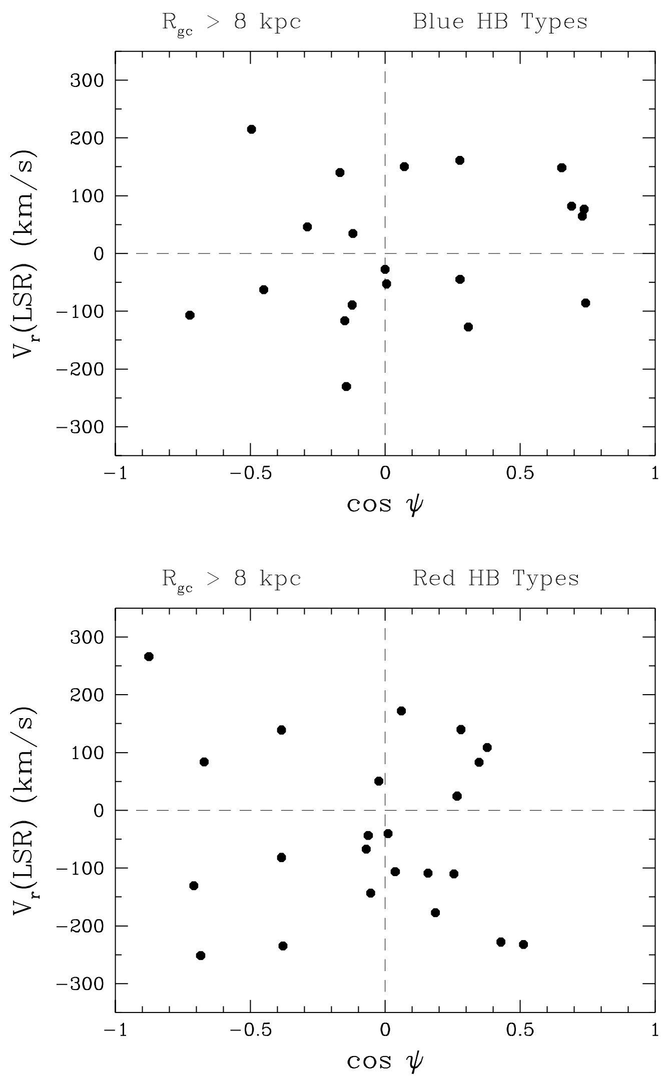

Fig. 1.27. *Upper panel:* Rotation solution for the outer-halo (8 − 40 kpc) clusters with blue HB types. *Lower panel:* Rotation solution for outer-halo clusters with red HB types; note NGC 3201 at upper left

clusters and 6 RHB clusters with $R_{gc} > 8$ kpc. For the BHB subset, the mean orbital eccentricity is $\langle e \rangle = 0.66 \pm 0.06$ and energy is $\langle E \rangle = -(5.2 \pm 0.9) \times 10^4$ km^2 s^{-2}. For the RHB subset, these numbers are $\langle e \rangle = 0.65 \pm 0.07$ and $\langle E \rangle = -(5.2 \pm 0.9) \times 10^4$ km^2 s^{-2}. In the BHB group, we find 5 prograde orbits, 3 retrograde, and 2 "plunging" types; in the RHB group, there are 3 prograde, 1 retrograde, and 2 plunging.

All these comparisons suggest to me that the two groups have no large collective differences in orbital properties; the rather modest differences in mean rotation speed are driven strongly by small-sample statistics. In addition, the normal assumption of an approximately isotropic orbital distribution for the halo clusters still seems to be quantitatively valid.

Where does this analysis leave the search for remnants of accreted satellites in the halo? My impression – perhaps a pessimistic one – is that distinct moving groups have proven almost impossible to find (if they exist in the first place) from the analyses of globular cluster motions. Once we start subdividing our meager total list of halo clusters by all the various parameters such as metallicity, spatial zones, or CMD morphology, the selected samples quickly become too small for statistically significant differences to emerge. The one outstanding exception is of course the four Sagittarius clusters, which are a physically close group that has not yet been tidally stretched out all around the halo. But even here, one suspects that they would not yet have been unambiguously realized to be part of a single system if their parent dwarf galaxy had not called attention to itself. If Sagittarius is a typical case of an accreted satellite, then we could reasonably expect that any others in the past would have brought in similarly small numbers of globular clusters – one, two, or a handful at a time – and thus extremely hard to connect long after the fact. My conclusion is that, if accreted remnant satellites in the halo are to be identified from this type of analysis, they will have to emerge from the study of halo field stars (e.g., Majewski et al. 1996), for which vastly larger and more statistically significant samples of points can be accumulated.

Three-dimensional space velocities would also, of course, be immensely valuable in the search for physically connected moving groups of clusters. Majewski (1994) notes that "the key test of common origin must come with orbital data derived from complete space velocities for these distant objects". Direct photometric searches for "star streams" trailing ahead of or behind disrupted satellites, have been proposed (see the reviews of Grillmair 1998; Johnson 1998). However, numerical experiments to simulate tidal stripping indicate that these disrupted streams would be thinly spread across the sky, enough so that they would be visible only for the largest satellites such as Sagittarius and the LMC.

1.3.5 Orbits in the Outermost Halo

Finding traces of disrupted satellites that are still connected along an orbital stream should be easiest in the outermost halo where the satellites originally resided and where the orbital timescales are the longest. It has been proposed several times, for example, that the Magellanic Clouds are connected along a great circle with other objects including the dwarf spheroidal satellites Draco, Ursa Minor, and Carina, and the small clusters Palomar 12 and Ruprecht 106 (Kunkel & Demers 1975, 1977; Lynden-Bell 1976). A similar stream comprising Fornax, Leo I and II, Sculptor, Palomar 3, Palomar 4, and AM-1 has been proposed (Majewski 1994; Lynden-Bell & Lynden-Bell 1995; Fusi Pecci et al. 1995). Correlation analyses of the radial velocities and locations of all the globular clusters and dwarf satellites have been carried out (Lynden-Bell & Lynden-Bell 1995; Fusi Pecci et al. 1995), with the result that several possible orbital "groups" have been proposed, usually containing just three or four clusters each. At this stage it is unclear how real any of these groups might be. (It is noteworthy, however, that this analysis successfully connected the Sagittarius clusters before the dwarf galaxy itself was found.)

From the viewpoint of space distribution alone, it is unquestionably true that the objects beyond $R_{gc} \sim 60$ kpc are not isotropically distributed around the Galaxy. Most of these objects do lie moderately close to a single plane (the Fornax-Leo-Sculptor stream) which is nearly perpendicular to the Galactic plane. (It should be kept in mind that their distribution is "planar" only in a relative sense; the thickness of the plane is about 50 kpc and the diameter about an order of magnitude larger.) This stream is shown in Fig. 1.28 (adapted from Majewski 1994), where we are looking at it in an orientation which minimizes the side-to-side spread in locations: we select a new axis $X' = X \cos\theta + Y \sin\theta$ where θ is the coordinate rotation angle between the X, X' axes and now X, Y are measured relative to the Galactic center. For the Fornax-Leo-Sculptor stream, Majewski (1994) finds $\theta \simeq 50°$. An alternate, and intriguing, interpretation of the strongly prolate distribution of these satellites is that they delineate the shape of the outermost dark matter halo of the Milky Way (see Hartwick 1996 for a kinematical analysis and complete discussion of this possibility).

1.3.6 Some Conclusions

A summary of the essential points that we have discussed in this section may be helpful.

- Mean rotation speeds and orbital velocity dispersions can be usefully estimated from cluster radial velocities and properly designed kinematics diagrams. The main sources of uncertainty in these plots are (a) uncertainties in the measured cluster distances, which are important for highly reddened clusters in the inner halo; and (b) the small numbers of points in any one subsample. Unfortunately, we can do nothing to improve our sample size of *clusters*, but the kinematics of the halo can also be studied through much larger samples of *field stars*.

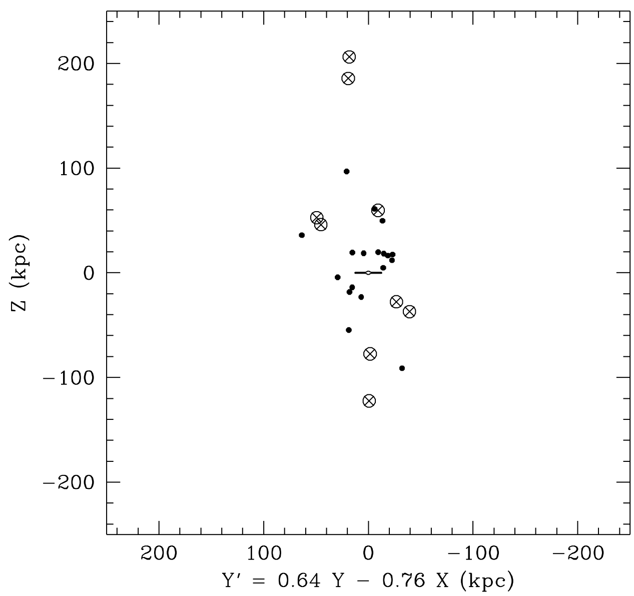

Fig. 1.28. Locations of remote satellites around the Milky Way: solid dots are halo clusters, crosses are dwarf galaxies. The X' axis is rotated $50°$ counterclockwise from the normal $X-$axis, so that we are looking nearly parallel to the proposed Fornax-Leo-Sculptor plane. Note that X, Y here are measured relative to the Galactic center

- The metal-rich (MRC) clusters have a strong systemic rotation with two weakly distinguishable subcomponents: an inner $(0-4$ kpc) bulge-like system with $V(rot) \sim 90$ km s^{-1}, and an outer $(4-8$ kpc) system with $V(rot) \sim 150$ km s^{-1} somewhat more like the thick disk.
- The most metal-poor (MPC) clusters, those with [Fe/H] $\lesssim -1.7$, have a systemic prograde rotation of $V(rot) \sim 80-100$ km s^{-1}, somewhat like the MRC bulge population.
- *Individual* clusters with retrograde orbits certainly exist, but:
- There are no unambiguous subgroups of clusters that have *systemic* retrograde rotation that are identifiable on the basis of Galactocentric distance, metallicity, or HB morphology. Previous suggestions of such retrograde groups seem to have arisen because of the unfortunate and uniquely strong influence of the single cluster NGC 3201 – again, a consequence of the small samples of objects and the effects of rare outliers on statistical distributions.
- To first order, the mid-to-outer halo MPC clusters can reasonably be described as forming a system with small net rotation and roughly isotropic orbit distribution. With the exception of the Sagittarius clusters, no "ac-

creted satellite" remnant groups have yet been reliably identified *from the clusters alone.* The main hope for identifying such groups lies with the analysis of field-star populations.

- Several of the outermost clusters ($R_{gc} > 50$ kpc) may be part of an extremely large-scale orbital stream (the Fornax-Leo-Sculptor stream).
- We urgently need more and better three-dimensional space motions for the halo clusters, measured through accurate absolute proper motions. Such information will allow us to determine the systemic rotation $V(rot)$ of the *outer* halo; the degree of anisotropy of the orbital distribution; the true fraction of retrograde orbits; and the identification of true orbital families and tidal streams.

1.4 THE MILKY WAY SYSTEM: DYNAMICS AND HALO MASS PROFILE

There is no illusion more dangerous than the belief that the progress of science is predictable.

Freeman Dyson

1.4.1 The Orbit Distribution

It is obvious from the large line-of-sight velocity dispersions quoted in the previous section that the globular clusters do not have circular or true disklike orbits as a group (though a few individual ones may). Neither do they have plunging, purely radial orbits as a group, since the large tangential motions of many of them are obvious. From the near-uniformity of the observed σ_{los} in any direction through the halo, we normally assume the orbital distribution to be *isotropic* (that is, $\sigma_U \simeq \sigma_V \simeq \sigma_W$ along any three coordinate axes). The best current evidence (Dinescu et al. 1999, from measurement of the three-dimensional space motions of the clusters) continues to support the isotropic assumption.

Another diagnostic of the cluster orbits which was used early in the subject (e.g., von Hoerner 1955; Kinman 1959b) and recently revived by van den Bergh (1993a,b) is the velocity ratio

$$u_0 = \frac{v_r(LSR)}{V(r)} \tag{1.23}$$

where $V(r)$ is the rotation speed for circular orbits at distance r from the Galactic center; and v_r is the radial velocity of the cluster relative to a stationary point at the Sun, as used in the previous Section. Now also let φ be the angle between Sun and Galactic center (GC), as seen from the cluster,

$$\cos\varphi = \frac{r - R_0 \cos b \cos\ell}{(r^2 + R_0^2 - 2rR_0 \cos b \cos\ell)^{1/2}} \,. \tag{1.24}$$

(see Fig. 1.29). If the cluster is on a circular orbit, then clearly $v_r(LSR) = V \sin\varphi$. However, if it is on a purely radial orbit with respect to the GC, then $v_r(LSR) = V \cos\varphi$, and if $u_0 \gtrsim \sqrt{2}$ for such cases then nominally the cluster would be on an escape orbit *if* there were no additional halo mass outside its current location. Figure 1.30 shows the distribution of ratios u_0 against φ for the clusters in two different areas of the halo, where we have assumed $V_0 \equiv 220$ km s^{-1} for all R_{gc}. Objects near the left-hand side of the diagram must be on strongly radial orbits, while ones on the right-hand side must be on more nearly circular orbits. The wide range of locations across the diagram confirms that neither purely radial nor circular orbits are dominant, and the fact that almost all the points stay comfortably below the upper envelope $u_0 = \sqrt{2}$ indicates – as it should – that they are well within the limits of bound orbits.

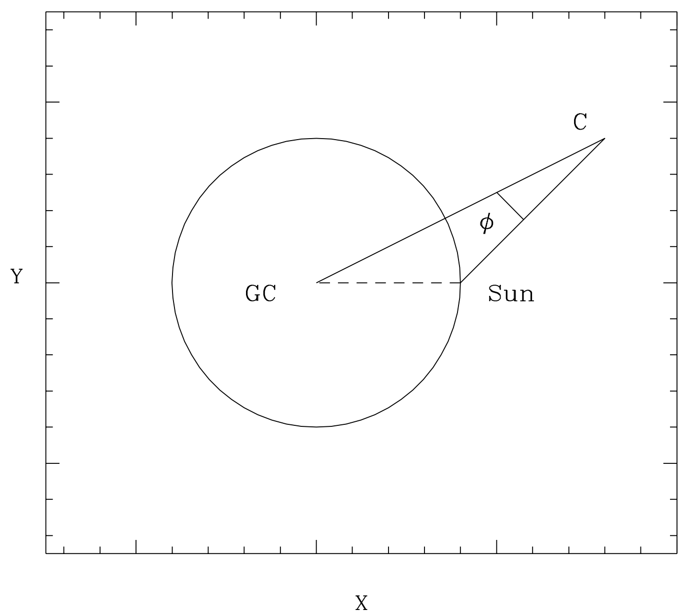

Fig. 1.29. Geometric definition of the angle φ between Sun and Galactic center (GC), as seen from the globular cluster C. The circle on the Galactic XY plane has radius R_0

1.4.2 The Mass of the Halo: Formalism

Early in the history of this subject, it was realized that the radial velocities of the globular clusters provided an extremely effective way to estimate the mass profile of the Galaxy out to large distances. Roughly speaking, the

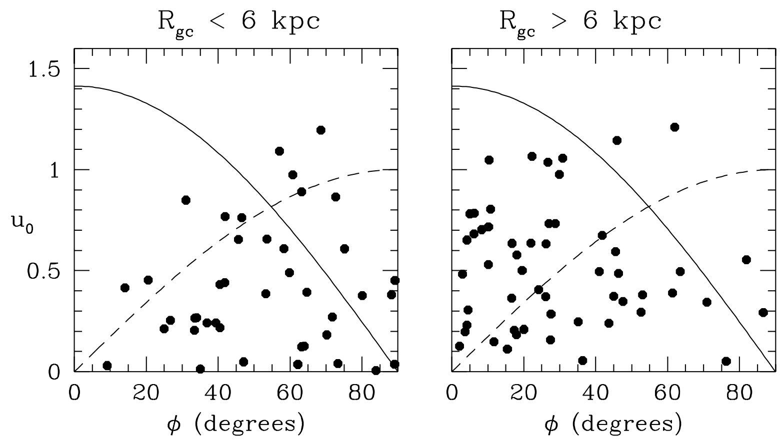

Fig. 1.30. Velocity-ratio diagram for cluster orbits as defined in the text. Any cluster on a purely circular orbit around the GC would lie on the dashed line $u_0 = \sin\varphi$. The solid line, $u_0 = 2^{1/2}\cos\varphi$, would be the locus of clusters on purely radial escape orbits if there were no additional halo mass outside their current location

velocity dispersion σ_v of a group of clusters at distance r from the Galactic center reflects the total mass $M(r)$ enclosed within r, $G\,M(r)/r \sim \sigma_v^2$. Knowing σ_v, we may then invert this statement and infer $M(r)$. The first comprehensive attempt to do this was by Kinman (1959b), with later and progressively more sophisticated analyses by (among others) Hartwick & Sargent (1978), Lynden-Bell et al. (1983), Little & Tremaine (1987), and Zaritsky et al. (1989). The limits to our knowledge of the halo mass profile are now set by our lack of knowledge of the true three-dimensional space motions of the clusters.

Interested readers should see Fig. 6 of Kinman (1959b), where the cluster velocity dispersions are used to plot the $M(r)$ profile in much the same way as is done here. Kinman's graph provides the earliest clear-cut evidence that I am aware of that $M(r)$ grows linearly out to at least $r \sim 20$ kpc in our Galaxy, thus showing the existence of its "dark-matter" halo – although it was not interpreted as such at the time. It was only 15 to 20 years later that astronomers routinely accepted the dominance of unseen matter in galaxies, although the key observations were in front of them, in basically correct form, far earlier.

Turning the first-order virial-theorem argument above into a quantitative formalism can be done in a variety of ways. Always, a major source of uncertainty is that we observe only one component of the true space velocity. Here, I will set up the formal analysis of the particle velocity distribution with an approach adapted from Hartwick & Sargent (1978); it has the advantage of physical clarity, and of displaying the key geometric parameters of the mass distribution in explicit form. Readers should see Little & Tremaine (1987) and Zaritsky et al. (1989) for alternate modern formulations based on Bayesian statistics.

Consider a system of particles orbiting in the halo with number density in phase space $f(x_i)$ with position and velocity coordinates x_i $(i = 1, \ldots, 6)$. The particles are imagined to be dominated by internal random motions (that is, they have a high radial velocity dispersion relative to the Galactic center, which provides the main support of the system against gravity), and the distribution is explicitly assumed time-independent, $\partial f / \partial t = 0$. At this point we also assume a roughly spherical mass distribution so that the potential energy is $U = U(r)$. We impose the appropriate physical boundary condition that $f \to 0$ as $x_i \to \infty$ for any coordinates or velocities. Finally, we adopt spherical polar coordinates (r, θ, φ) and velocity components (R, Θ, Φ).

The collisionless Boltzmann equation for such a system is then

$$\frac{df}{dt} = 0 = \dot{r}\frac{\partial f}{\partial r} + \dot{R}\frac{\partial f}{\partial R} + \dot{\Theta}\frac{\partial f}{\partial \Theta} + \dot{\Phi}\frac{\partial f}{\partial \Phi} \tag{1.25}$$

(see, for example, Chp. 4 of Binney & Tremaine 1987), and where we have already dropped any derivatives $\partial/\partial\theta, \partial/\partial\varphi$ from spherical symmetry. We can write the various force components as derivatives of the potential energy,

$$F_r = \frac{-\partial U}{\partial r} = \dot{R} - \frac{\Theta^2}{r} - \frac{\Phi^2}{r} \tag{1.26}$$

$$F_\theta = 0 = -\frac{1}{r}\frac{\partial U}{\partial \theta} = \dot{\Theta} + \frac{R\Theta}{r} - \frac{\Phi^2}{r}\cot\,\theta \tag{1.27}$$

$$F_\varphi = 0 = -\frac{1}{r\sin\theta}\frac{\partial U}{\partial \varphi} = \dot{\Phi} + \frac{R\Phi}{r} + \frac{\Theta\Phi}{r}\cot\,\theta \tag{1.28}$$

When we use these to substitute expressions in for $\dot{R}, \dot{\Theta}, \dot{\Phi}$, the Boltzmann equation becomes

$$\begin{aligned} 0 \;=\; & R\frac{\partial f}{\partial r} + \Big(\frac{\Theta^2 + \Phi^2}{r} - \frac{dU}{dr}\Big)\frac{\partial f}{\partial R} + \frac{1}{r}\Big(\Phi^2 \cot\,\theta \;-\; R\;\Theta\Big)\frac{\partial f}{\partial \Theta} \\ & - \frac{1}{r}\big(R\Phi \;+\; \Theta\;\Phi\;\cot\,\theta\big)\frac{\partial f}{\partial \Phi} \;\equiv\; g(r, R, \Theta, \Phi) \end{aligned} \tag{1.29}$$

Next take the first $R-$moment of this equation and integrate over all velocities:

$$0 = \int\limits_R \int\limits_\Theta \int\limits_\Phi g \;\cdot\; R\; dR\; d\Theta\; d\Phi \tag{1.30}$$

and also define the number density $\nu(r)$ (number of particles per unit volume of space) as

$$\nu \;\equiv\; \int\limits_R \int\limits_\Theta \int\limits_\Phi f \cdot dR\, d\Theta\, d\Phi\,, \tag{1.31}$$

and in general define the mean of any quantity Q as $\langle Q \rangle \equiv \frac{1}{\nu} \int Q\, f\, dR\, d\Theta\, d\Phi$. When we carry out the integration in Eqn. 1.30, we find that several of the terms conveniently vanish either from symmetry or from our adopted boundary conditions (we will leave this as a valuable exercise for the reader!), and the remainder reduces to

$$0 = \frac{d}{dr}\big(\nu\langle R^2\rangle\big) - \frac{\nu}{r}\langle\Theta^2 + \Phi^2\rangle + \nu\frac{dU}{dr} \; + \; \frac{2\,\nu}{r}\langle R^2\rangle\,. \tag{1.32}$$

Now we can put in the radial force component

$$\frac{dU}{dr} \; = \; \frac{G\;M(r)}{r^2} \tag{1.33}$$

where $M(r)$ is the mass contained within radius r, and we reduce (1.32) further to

$$G\;M(r) \; = \; r\;\langle R^2\rangle\left(\frac{\langle\Theta^2 + \Phi^2\rangle}{\langle R^2\rangle} \; - \; \frac{r}{\nu}\frac{d\nu}{dr} \; - \; \frac{r}{\langle R^2\rangle}\frac{d\langle R^2\rangle}{dr} \; - \; 2\right). \tag{1.34}$$

The second and third terms in the brackets of (1.34) can be simplified for notation purposes if we denote them as α and β, such that the density of the GCS is assumed to vary with radius as $\nu \sim r^{\alpha}$, and the radial velocity dispersion as $\langle R^2\rangle \sim r^{\beta}$. We can also recognize that the halo has some net rotation speed Θ_0 such that the tangential velocity can be broken into rotational and peculiar (random) parts,

$$\Theta \; = \; \Theta_0 \; + \; \Theta_p \tag{1.35}$$

$$\langle\Theta^2\rangle \; = \; \Theta_0^2 \; + \; \langle\Theta_p^2\rangle \tag{1.36}$$

and finally for notation purposes we define the *anisotropy parameter*

$$\lambda \; \equiv \; \frac{\langle\Theta_p^2 + \Phi^2\rangle}{\langle R^2\rangle}\,. \tag{1.37}$$

For purely radial orbits, clearly $\lambda \rightarrow 0$, and for isotropically distributed orbits where all three random velocity components are similar, $\lambda = 2$. With these various abbreviations, the simplified moment of the Boltzmann equation takes the final form that we need,

$$G\;M(r) \; = \; \langle r\rangle\,\langle R^2\rangle\Big(\lambda - \alpha - \beta + \frac{\Theta_0^2}{\langle R^2\rangle} - 2\Big)\,. \tag{1.38}$$

The quantities on the right-hand side represent the characteristics of our tracer population of particles (the globular clusters), while the single quantity on the left ($M(r)$) represents the gravitational potential to which they are reacting. This form of the equation displays rather transparently how the deduced mass $M(r)$ depends on the radial velocity dispersion $\langle R^2\rangle$ of a set of points with mean Galactocentric radius $\langle r\rangle$; on the orbit anisotropy λ and the rotation speed Θ_0; and on the density distribution ν and radial dependence of the velocities β. It is easier to achieve a given radial velocity dispersion if the cluster orbits are more purely radial (smaller λ, thus smaller enclosed mass M). Similarly, a steeper halo profile (more negative α or β) or a larger systemic rotation speed Θ_0 will lead to a larger mass for a given velocity dispersion.

Can we simplify this relation any further? On empirical grounds we expect from our accumulated evidence about the Galaxy to find roughly isotropic orbits $\lambda \sim 2$, a nearly isothermal halo and thus constant velocity dispersion $\beta \sim 0$, and a small halo rotation $\Theta_0^2 \ll \langle R^2 \rangle$. Thus as a first-order guess at the mass distribution we might expect very crudely

$$G\ M(r) \simeq -\alpha \langle r \rangle \langle R^2 \rangle \,. \tag{1.39}$$

At this point, we could now select groups of clusters at nearly the same radial range $\langle r \rangle$, calculate their velocity dispersion $\langle R^2 \rangle$, and immediately deduce the enclosed mass $M(r)$. We will see below, however, that it will be possible to refine this guess and to place somewhat better constraints on the various parameters.

1.4.3 The Mass of the Halo: Results

To calculate the mass profile $M(r)$ for the Milky Way, we will now use the radial velocity data for all the MPC clusters ([Fe/H] < -0.95), which form a slowly rotating system dominated by internal random motions. We will, of course, find that we rapidly run out of clusters at large r, just where we are most interested in the mass distribution. To help gain statistical weight, we will therefore add in the data for nine satellites of the Milky Way, recognizing fully (see the previous Section) that the orbital motions of some of these may well be correlated. These nine include the Magellanic pair (LMC+SMC) plus the dwarf spheroidals more remote than ~ 50 kpc from the Galactic center (Draco, Ursa Minor, Carina, Sextans, Sculptor, Fornax, Leo I and II). Furthermore, we will also put in data for 11 remote RR Lyrae and horizontal-branch stars in the halo in the distance range 40 kpc $< r <$ 65 kpc, as listed by Norris & Hawkins (1991). To keep track of the possibility that these latter halo *field stars* might have (say) a different anisotropy parameter λ than the clusters, we will keep them in a separate bin from the clusters.

The relevant data for the dwarf satellites are summarized in Table 1.7. Successive columns list the Galactic latitude and longitude in degrees; the measured horizontal-branch magnitude V_{HB} of the old-halo or RR Lyrae stars; the foreground reddening; the intrinsic distance modulus and mean metallicity of the stars; and the heliocentric radial velocity (km s^{-1}). For all but the LMC and Leo I, the distance is calculated from V_{HB} and our adopted prescription $M_V(HB) = 0.15$[Fe/H] + 0.80. For Leo I, the distance estimate relies on the $I-$band magnitude of the RGB tip.

The velocity dispersion data and calculated mass profile are listed in Table 1.8. Here, the MPC clusters have been divided into radial bins with roughly a

Table 1.7. Dwarf satellites of the Milky Way

Satellite	ℓ	b	V_{HB}	E_{B-V}	μ_0	[Fe/H]	v_r	Sources
LMC	280.47	−32.89	19.20	0.06	18.54		245.0	1,2
UrsaMinor	104.95	44.80	19.90	0.01	19.37	−2.2	−247.4	3,4
Sculptor	287.54	−83.16	20.10	0.03	19.46	−1.8	109.9	5,6
Draco	86.36	34.71	20.10	0.02	19.54	−2.1	−293.3	4,7
Sextans	243.50	42.27	20.36	0.03	19.75	−2.05	227.9	8,9
Carina	260.11	−22.22	20.65	0.03	19.96	−1.52	223.1	10,11
Fornax	237.10	−65.65	21.25	0.02	20.64	−1.4	53.0	12,13
Leo II	220.17	67.23	22.18	0.02	21.52	−1.9	76.0	14,15
Leo I	225.98	49.11	22.75:	0.02	22.18	−2.0	285.0	16,17

Sources: (1) This paper (Section 2) (2) NASA Extragalactic Database (NED) (3) Nemec et al. 1988 (4) Armandroff et al. 1995 (5) Da Costa 1984 (6) Queloz et al. 1995 (7) Carney & Seitzer 1986 (8) Suntzeff et al. 1993 (9) Mateo et al. 1995 (10) Smecker-Hane et al. 1994 (11) Mateo et al. 1993 (12) Smith et al. 1996 (13) Mateo et al. 1991 (14) Mighell & Rich 1996 (15) Vogt et al. 1995 (16) Lee et al. 1993b (17) Zaritsky et al. 1989

dozen objects per bin; successive columns list the mean Galactocentric radius, number of clusters in the bin, mean radial velocity dispersion, and enclosed mass determined as described below.

Table 1.8. Radial velocity dispersion profile

$\langle R_{gc}\rangle$ (kpc)	n	$\langle R^2\rangle^{1/2}$	$M(r)$ $(10^{11} M_\odot)$	Comment
1.33 ± 0.12	14	133 ± 24	0.17 ± 0.03	
2.38 ± 0.10	8	136 ± 32	0.31 ± 0.08	
3.83 ± 0.14	12	99 ± 19	0.30 ± 0.06	
5.55 ± 0.20	9	80 ± 18	0.31 ± 0.07	
7.55 ± 0.26	10	127 ± 27	0.88 ± 0.19	
10.4 ± 0.28	9	145 ± 33	1.40 ± 0.32	
16.3 ± 0.71	15	135 ± 24	1.88 ± 0.34	
29.2 ± 2.1	10	141 ± 30	3.93 ± 0.88	
53.8 ± 2.5	11	112 ± 22	4.94 ± 1.03	RR Lyraes
95.1 ± 8.1	6	109 ± 30	8.97 ± 2.61	Outer GCs only
71.0 ± 6.2	5	113 ± 34	6.95 ± 2.22	GCs + dSph
131.0 ± 20.7	10	87 ± 19	8.28 ± 2.19	GCs + dSph

In Fig. 1.31, we first show the r.m.s. velocity dispersion $\langle R^2\rangle^{1/2}$ plotted against Galactocentric distance. The velocities R are the radial velocities of the clusters after removal of the LSR motion V_0, i.e. they are the same as σ_{los} used in the discussion of kinematics in Section 3. From the graph, we see that σ stays nearly uniform at $\sim$ 130 km s^{-1} for $r \lesssim 50$ kpc and then declines gradually at larger distances. An extremely crude first-order model for this behavior would be a simple isothermal halo for which the velocity dispersion $\sigma \sim$ *const* and $M(r)$ increases in direct proportion to r, out to some truncation limit r_h beyond which the density is arbitrarily zero. At larger distances the velocity dispersion would then decline as $\sigma \sim r^{-1/2}$. For a choice of halo 'limiting radius' $r_h \simeq 60$ kpc, this oversimplified model actually represents the data points quite tolerably. Whether or not we exclude the field-star point (labelled RR) makes little difference, suggesting that there are probably no gross intrinsic differences between the clusters and field stars.

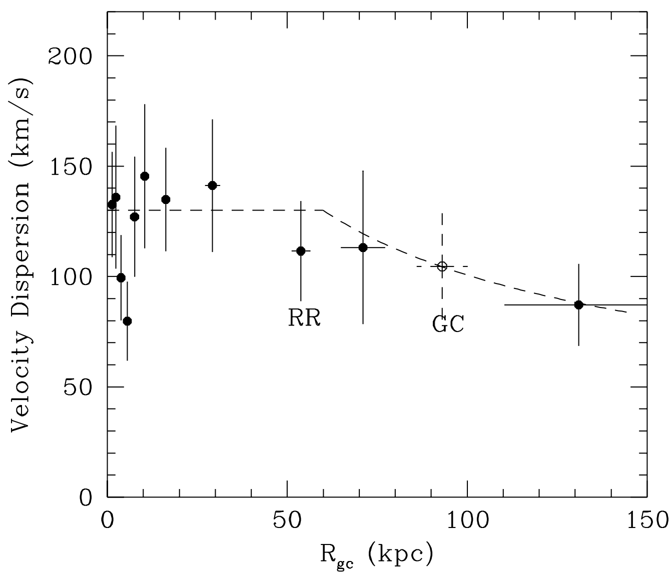

Fig. 1.31. Radial velocity dispersion (line-of-sight relative to a fixed point at the Sun) plotted against Galactocentric distance. The point labelled 'RR' is for the 11 field RR Lyrae and HB stars described in the text. The point labelled 'GC' is for the six most remote clusters without including any of the dwarf satellite galaxies. The *dashed line* shows the expected trend for the velocity dispersion if the halo were ideally isothermal out to a radius $r_h \sim 60$ kpc, and then truncated abruptly there.

For our final calculation of the mass profile $M(r)$, we employ Eqn. 1.38 and add in what we know about the space distribution and kinematics of our test particles, the globular clusters:

- From Fig. 1.3, we see that the space density exponent α steepens smoothly from an inner-halo value of $\simeq -2$ up to -3.5 in the outer halo.

- For the anisotropy parameter, we have $\lambda = 1.2 \pm 0.3$, determined directly from the three-dimensional space motions summarized by Dinescu et al. (1999). This value represents a mild anisotropy biased in the radial direction, and appears not to differ significantly with location in the halo, at least for the metal-poor clusters that we are using.
- The rotation velocity Θ_0 is $\simeq 80$ km s^{-1} for $R \lesssim 8$ kpc and $\simeq 20$ km s^{-1} for $R \gtrsim 8$ kpc (see Section 3 above).
- Lastly, we adopt $\beta \simeq 0$ since no strong variation of velocity dispersion with radius is evident.

The resulting mass profile is shown in Fig. 1.32. As expected, $M(r)$ grows linearly for $r \lesssim 40$ kpc but then increases less steeply, reaching $\sim (8 \pm 2) \times 10^{11} M_{\odot}$ at $R \sim 100$ kpc.

An extra useful comparison – at least for the inner part of the halo – is our knowledge of the rotation speed of the Galactic *disk*, $V_0 = 220$ km s^{-1} roughly independent of radius. Within the region of disk/halo overlap, we can then write

$$G\ M\ =\ r\, V_0^2 \tag{1.40}$$

which becomes $M(r) = 1.13 \times 10^{10} M_{\odot}\ r(\mathrm{kpc})$. This relation matches the GCS dispersion data in Fig. 1.32 extremely well for $r \lesssim 50$ kpc.

With the foregoing arguments, we have been able to extend the mass profile for the Milky Way outward to about three times further than in Kinman's original attempt 40 years ago. However, our suggestion that the halo density begins to "die" somewhere around $r_h \sim 60$ kpc is risky, since we do not know if our model assumptions apply to the outermost objects. The motions of several of these clusters and dwarf satellites may be *correlated* (see Section 3), which would invalidate our adopted values for λ, Θ_0, and $\langle R^2 \rangle$. Perhaps equally important is that the most remote single object in our list, Leo I, may not be bound to the Galaxy (see Lee et al. 1993b; Zaritsky et al. 1989). Of the dozen most remote objects, Leo I has both the largest distance ($R_{gc} = 277$ kpc) and largest velocity ($v_r = 175$ km s^{-1}) and thus carries significant weight in the solution by itself. There is no way to rule out the likely possibility that it is simply a dwarf moving freely within the Local Group, like others at similar distances from M31. To *marginally* bind Leo I to the Milky Way would require $M(\mathrm{total}) \gtrsim 10^{12} M_{\odot}$ within 100 kpc, which in turn would be marginally inconsistent with the mass given by all the other remote satellites *if* their motions are independent and roughly isotropic.

A widely used analytic model for dark-matter halos with some basis in cosmological N-body simulations is that of Navarro et al. (1996; denoted NFW), giving a density profile

$$\rho(r)\ =\ \frac{\rho_0}{\frac{r}{r_s}\left(1 + \frac{r}{r_s}\right)^2} \tag{1.41}$$

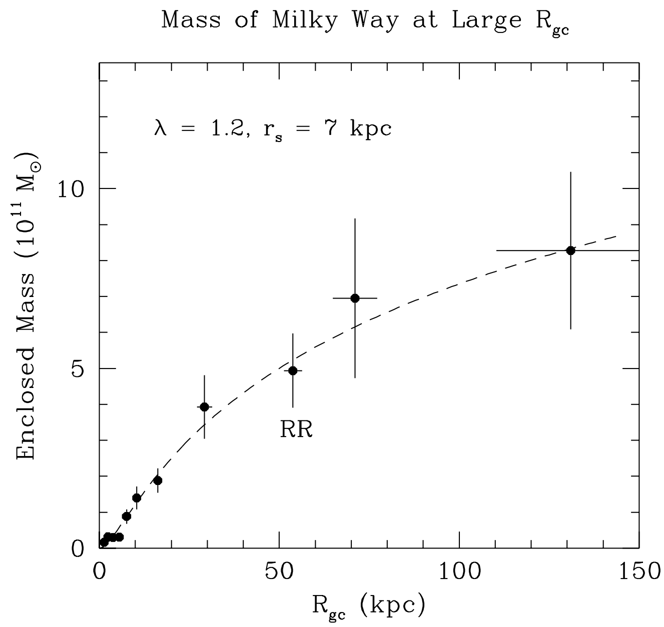

Fig. 1.32. Mass profile for the Milky Way halo, with the "best fit" set of parameters discussed in the text. The dashed line is an NFW model halo for a characteristic radius $r_s = 7$ kpc

where the free parameter r_s is a scale radius chosen to fit the galaxy concerned. Integration of this spherically symmetric profile yields for the enclosed mass

$$M(r) = \text{const} \left(\ln \left(1 + \frac{r}{r_s} \right) - \frac{(r/r_s)}{\left(1 + \frac{r}{r_s}\right)} \right) \tag{1.42}$$

An illustrative mass profile for this model is shown in Fig. 1.32 as the dashed line; for the Milky Way halo, a suitable scale radius is evidently near $r_s \sim 7$ kpc. At large radius, the NFW model profile falls off as $\rho \sim r^{-3}$ and the enclosed mass $M(r)$ keeps growing logarithmically with r. Whether or not the real Milky Way halo genuinely agrees with this trend, or whether there is a steeper cutoff in the density profile past $R_{gc} \sim 60$ kpc, is impossible to say at present. Some guidance may be obtained from measurements of the halo surface density of M31 along its minor axis by Pritchet & van den Bergh (1994), who find that the halo light drops more steeply (more like $\rho \sim r^{-5}$ at large projected radius) than the NFW model. If the same is true for the Milky Way, then we might anticipate that the halo mass 'converges' somewhere near $8 \times 10^{11} M_\odot$.

A promising avenue to determining the total mass of the Galaxy more accurately would be to obtain true three-dimensional space motions for the outermost satellites. Absolute proper motions have been obtained for the Magellanic Clouds (e.g., Kroupa & Bastian 1997a,b; Jones et al. 1994), leading to a total mass estimate for the Galaxy (Lin et al. 1995) of $\sim 5.5 \times 10^{11} M_{\odot}$ within 100 kpc, consistent with what we have derived here. However, similar data for several of the much more distant ones will be needed for a reliable answer to emerge.

This rather frustrating state of uncertainty is where we will have to leave the subject.

1.5 THE MILKY WAY SYSTEM: THE LUMINOSITY FUNCTION AND MASS DISTRIBUTION

Generally, researchers don't shoot directly for a grand goal ... [but] our piecemeal efforts are worthwhile only insofar as they are steps toward some fundamental question.

Martin Rees

The *integrated luminosity* of a globular cluster is the cumulative light of all its stars, and as such it is a direct indicator of the total cluster mass as long as we know the total mass-to-light ratio. We will see later (Section 8) that the distribution of cluster masses is a major clue to understanding their process of formation. In this section, we will investigate what this distribution looks like, and see how it encouraged the first outward steps to comparisons of globular cluster systems in other galaxies.

1.5.1 Defining the GCLF

For a whole ensemble of globular clusters, the relative number *per unit magnitude* is called the globular cluster luminosity function or GCLF.[9] This distribution is plotted in Fig. 1.33 for 121 Milky Way clusters with moderately reliable data (Harris 1996a; of 140 clusters with measured total magnitudes and distances, we reject 19 with extremely high reddenings). Formally, the total cluster magnitude M_V^T is independent of foreground absorption, as it is simply the difference between the apparent total magnitude and apparent distance modulus, $M_V^T = V^T - (m-M)_V$. However, the quality of both V^T and the measured distance obviously degrades with increased reddening, for the reasons discussed in Section 1. We note in passing that the zeropoint of the luminosity scale varies directly with the adopted zeropoint of the RR Lyrae distance scale ($M_V(\mathrm{HB}) = 0.50$ at $[\mathrm{Fe/H}] = -2.0$; see Section 2), but the overall shape of the distribution is nearly independent of the [Fe/H]

[9] Do not confuse the GCLF with the number of stars per unit magnitude within one cluster, which is the *stellar* luminosity function.

coefficient of the distance scale since clusters of all metallicities lie at all luminosities. For example, arbitrarily adopting $\alpha = 0$ for the slope of the RR Lyrae luminosity calibration would change the peak point of the GCLF by less than 0.05 mag and have negligible effect on the standard deviation.

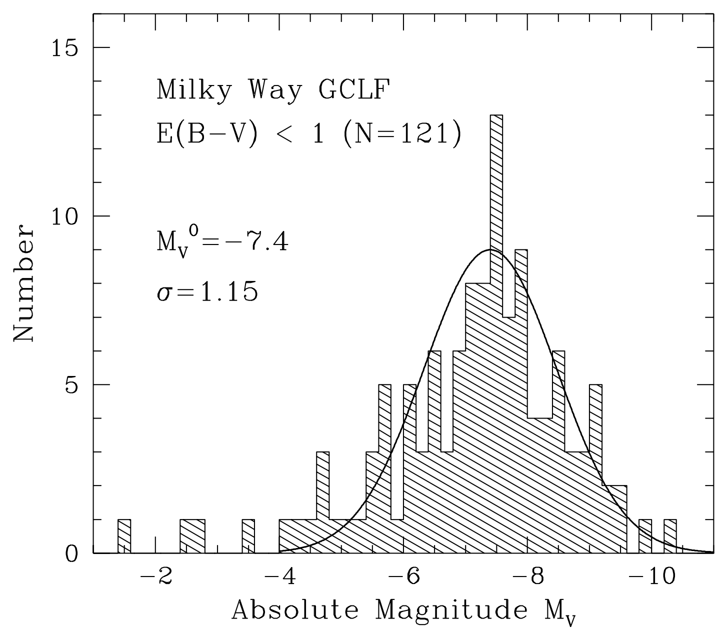

Fig. 1.33. Number of globular clusters per 0.2-magnitude bin, for the Milky Way. A Gaussian curve with mean $M_V = -7.4$ and standard deviation $\sigma = 1.15$ mag is superimposed to indicate the degree of symmetry of the distribution

The GCLF is strikingly simple: unimodal, nearly symmetric, and rather close to a classic Gaussian shape. Very luminous clusters are rare, but (perhaps counterintuitively) so are faintest ones. The GCLF that we see today *must* be a combined result of the initial mass spectrum of cluster formation, and the subsequent $\gtrsim 12$ Gigayears of dynamical evolution of all the clusters within their parent galaxy. Do the features of the distribution depend in any obvious way on other observable quantities, or on which galaxy they are in? Is it a "universal" function? There are two immediate reasons for raising these questions: the first, which we will take up again in Section 7, is to use the GCLF as a standard candle for extragalactic distance determination. The second and more astrophysically important reason, which we will explore further in Sections 8 and 9, is to investigate how globular clusters form.

1.5.2 Correlations Within the Milky Way

An obvious starting point for the Milky Way clusters is to compare the GCLFs for the two major subpopulations (MRC and MPC). These are shown in Fig. 1.34. At first glance, the MRC distribution is the broader of the two, favoring fainter clusters a bit more once we take into account the different total numbers. (An earlier comparison with less complete data is made by Armandroff 1989.) But an obvious problem with such a statement is that we are comparing groups of clusters with rather different *spatial distributions*: all the MRC clusters are close to the Galactic center, but roughly half the MPC clusters are in the mid-to-outer halo, where they have had the luxury to be relatively free of dynamical erosion due to bulge shocking, dynamical friction, or even disk shocking. If we compare clusters of both types in the same radial zone, we obtain something like what is shown in Fig. 1.35. Excluding the thinly populated low-luminosity tail of the distribution ($M_V^T > -6$) and using only the clusters within $R_{gc} < 8$ kpc, we find that the GCLFs are virtually identical. In short, we have little reason to believe that cluster luminosity depends strongly on metallicity.

We have a much better *a priori* reason to expect the GCLF to depend on Galactocentric distance, because the known processes of dynamical erosion depend fairly sensitively on the strength of the Galactic tidal field. Thus in an idealized situation where all clusters were born alike, they would now have masses (luminosities) that were clear functions of location after enduring a Hubble time's worth of dynamical shaking. Other factors will, of course, enter to confuse the issue (the degree of orbital anisotropy might also change with mean R_{gc} and thus modify the shocking rates; and in the absence of any quantitative theory of cluster formation, we might also imagine that the mass spectrum at formation could depend on location within the protoGalaxy).

The distribution of luminosity M_V^T with location is shown in Fig. 1.36. Two features of the distribution are apparent to eye inspection: (a) the average luminosity does change with location, rising gradually to a peak somewhere around log $R_{gc} \sim 0.9$ and then declining again further outward; and (b) there is a progressive outward "spreading" of the distribution over a larger M_V^T range. These trends are roughly quantified in Table 1.9, where the mean and standard deviation of the GCLF are listed for six rather arbitrarily selected radial bins. The last line gives the resulting parameters for the entire combined sample excluding only the faintest few clusters ($M_V^T > -4.5$; cf. the Gaussian curve with these parameters in Fig. 1.33). Other recent discussions of the observations are given by Kavelaars & Hanes (1997), Gnedin (1997), and Ostriker & Gnedin (1997), who also find radial trends in the GCLF peak at the $\pm 0.2 - 0.3$ magnitude level and assert that these are likely to be due to differences in their rates of dynamical evolution (assuming, of course, a similar initial mass spectrum).

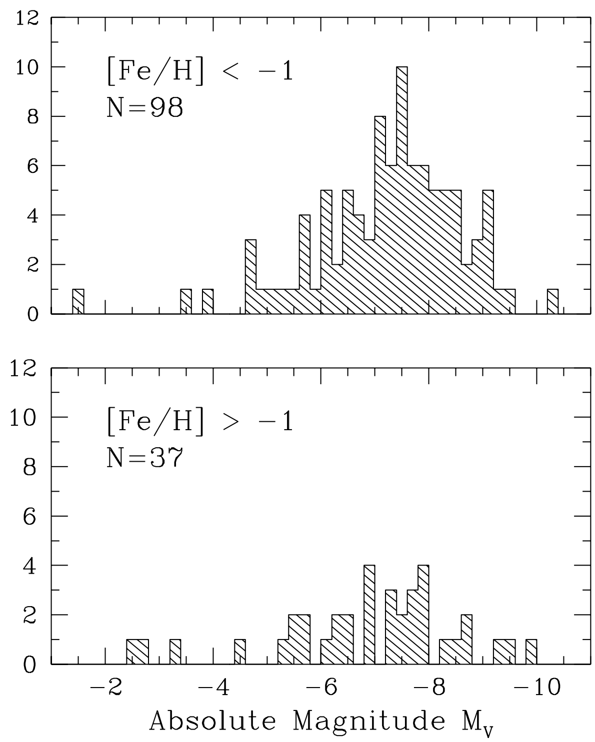

Fig. 1.34. GCLFs for the metal-poor clusters (upper panel) and metal-rich clusters (lower panel) in the Milky Way. There are five additional clusters not shown here which have measured absolute magnitudes, but unknown metallicities

Table 1.9. GCLF parameters versus Galactocentric distance

R_{gc} Range	N	$\langle M_V^T \rangle$	$\sigma(M_V)$
0 – 2 kpc	26	-7.28 ± 0.18	0.93 ± 0.12
2 – 5 kpc	40	-7.41 ± 0.17	1.08 ± 0.12
5 – 8 kpc	20	-7.46 ± 0.27	1.21 ± 0.18
8 – 15 kpc	17	-7.97 ± 0.22	0.91 ± 0.15
15 – 42 kpc	22	-7.06 ± 0.24	1.11 ± 0.16
> 60 kpc	6	-5.91 ± 0.76	1.86 ± 0.52
All R_{gc}	131	-7.40 ± 0.11	1.15 ± 0.08

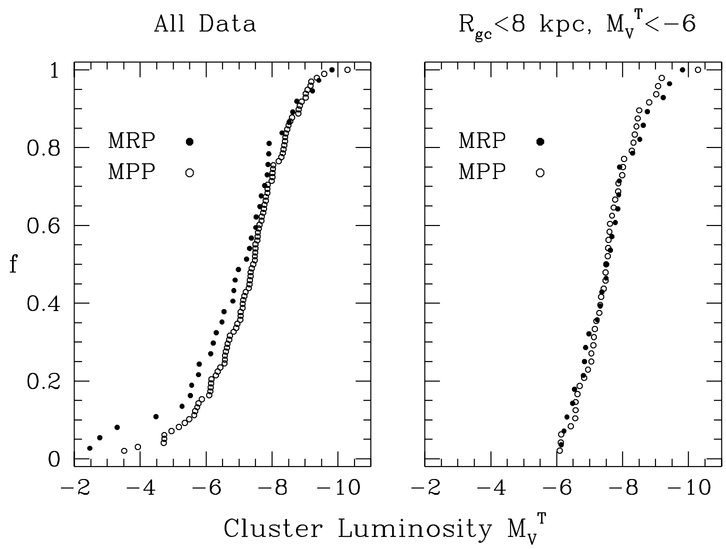

Fig. 1.35. Cumulative luminosity distributions for Milky Way globular clusters, separated by metallicity. The metal-rich population (here labelled MRP) is shown as the solid dots, and the metal-poor population (MPP) as open circles. The fraction of the total in each group is plotted against M_V^T. *Left panel:* all clusters with known luminosities are included. The difference between the two distributions is mainly in the low-luminosity tail. *Right panel:* Clusters within 8 kpc of the Galactic center and more luminous than $M_V = -6$. No difference is apparent

1.5.3 Dynamical Effects

Any star cluster has a limited lifetime. Even if it is left in isolation, the slow relaxation process of star-star encounters will eject individual stars, while the more massive stars (and binaries) will sink inward in the potential well, driving a net energy flow outward and a steady increase in central concentration. The existence of a tidal cutoff imposed by the Galaxy will serve to enhance the process of stellar evaporation, and move the cluster more rapidly toward core collapse and eventual dissolution. An early phase of cluster evolution that is not yet well understood is its first $\sim 10^7 - 10^8$ y, during which its initial population of high-mass stars evolves. In this stage, the residual gas left after cluster formation as well as most of the material in the massive stars will be ejected through stellar winds and supernova, thus expanding the cluster, enhancing tidal losses, and (possibly) leading it down the road to rapid disruption.

Closer in to the Galactic center, other mechanisms will become much more important. If a cluster is on a high-eccentricity orbit, the impulsive shock received near perigalactic passage will pump energy into its stellar

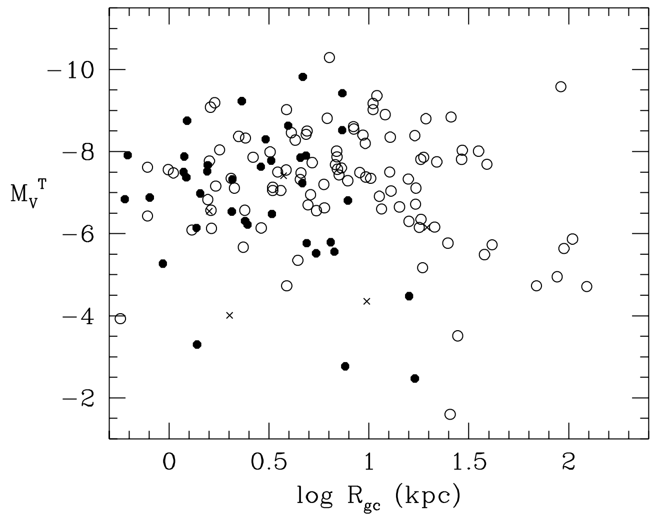

Fig. 1.36. Globular cluster luminosity versus location in the Galaxy, plotted as absolute integrated magnitude M_V^T versus log R_{gc}. MRC clusters are *filled symbols*; MPC clusters are *open symbols*; and clusters with unknown metallicities are *crosses*

orbits, hastening the overall expansion of the outer envelope of the cluster and thus its loss of stars to the field. In a disk galaxy, clusters on high-inclination orbits will feel similar impulsive shocks, with similar destructive results; while clusters that stay completely within the disk may encounter shocks from giant molecular clouds (GMCs) that have masses similar to theirs. Lastly, if a cluster already relatively close to the galactic center finds itself unluckily in a rotating bar-like potential, it can be forced into a chaotic orbit which can at some point take it right through the nucleus of the galaxy and thus dissolve it at a single stroke (Long et al. 1992). Clusters that are less massive, or structurally more diffuse, are more subject to these strong kinds of tidal disruptions.

Very close in to the galactic center, the classic effect of dynamical friction also takes hold (Tremaine et al. 1975). As the cluster moves through the galactic bulge, the field stars are drawn towards it as it passes through. A slight density enhancement of stars is created just behind the cluster, and this acts as a slow gravitational brake on its motion, causing the cluster orbit to spiral in to eventual destruction in the nucleus. This effect works faster on more massive clusters, and it is likely that any clusters (or dwarf galaxies) with $M \gtrsim 10^8 M_\odot$ would not have survived if they passed anywhere within the disk or bulge of the Galaxy (cf. the references cited below).

The long list of erosive mechanisms makes the galaxy sound like a dangerous place indeed for a star cluster to live. Clearly all of these processes are operating simultaneously, but the strength and rapidity of each of them depends strongly on the cluster's location within the halo or disk, its orbital eccentricity, and its mass and initial structure. The overall concept that the clusters we see today are the hardiest "survivors" of some larger original population was introduced by Fall & Rees (1977) and has dominated the thinking in this subject since then. Ideally, we would like to perform a comprehensive numerical simulation of an entire GCS in which *all* the various mechanisms are treated realistically over 10^{10} y, and thus find out what types of initial cluster distributions lead to GCSs like the one we see today. This is still a formidable task, not just because of sheer demands on computing power, but also because the initial conditions are poorly known. (What was the original distribution of cluster masses and structures? How important are possible differences in the stellar mass function (IMF) at formation? What was the initial space distribution of clusters like, and what about their starting distribution of orbits?) The potential range of parameter space is huge, and must eventually be linked to a complete theory of cluster formation.

Nevertheless, steady progress has been made, to the point where several excellent attempts at evolutionary syntheses have now been made. Recent examples include Capriotti & Hawley (1996); Gnedin & Ostriker (1997); Murali & Weinberg (1997a,b,c); Vesperini (1997, 1998); Vesperini & Heggie (1997); Baumgardt (1998); and Okazaki and Tosa (1995), among others. These papers are extremely valuable for illuminating the relative importance of the various mechanisms at different places in the host galaxy.

- At short distances ($R_{gc} \lesssim 1-2$ kpc), dynamical friction will remove the most massive clusters. At larger distances, dynamical friction becomes negligible for the typical globular clusters we see today.
- In the bulge and inner halo regions ($R_{gc} \lesssim 6$ kpc), bulge shocking and disk shocking are the dominant effects, acting particularly to destroy lower-mass clusters ($M \lesssim 10^5 M_\odot$) and to partially erode higher-mass ones.
- At still larger distances out in the halo, all the processes generally weaken to timescales longer than a Hubble time, and the slow mechanism of tidal evaporation becomes, somewhat by default, the dominant one.

How much of the current GCLF shape has been *produced* by sheer dynamical evolution is still unclear. It is reasonable to suspect that many more clusters at low masses must have existed initially, but that the higher-mass ones have been more immune to removal. Interestingly, an initial power-law mass distribution function such as $dN/dM \sim M^{-2}$, with many low-mass objects, evolves quickly into something resembling the traditional Gaussian in ($dN/d \log M$), and once the GCLF acquires this symmetric Gaussian-like form, it seems able to maintain that form for long periods (cf. Vesperini 1998; Murali & Weinberg 1997c; Baumgardt 1998). Most of the lowest-mass objects ($M \lesssim 10^4 M_\odot$) are disrupted even if they were initially present in large numbers. At higher masses, individual clusters evolve to lower mass and thus slide downward through the GCLF, rather slowly for high-mass clusters and more quickly for the smaller objects at the low-mass tail. But the shape of the overall distribution and the critical parameters (turnover point and dispersion) change rather slowly. Much more work remains to be done with different initial conditions to see how robust these first predictions are.

1.5.4 Analytic Forms of the GCLF

Since the GCLF was first defined observationally, a variety of simple functions have been applied to it for interpolation purposes. By far the most well known of these is the Gaussian in number versus magnitude,

$$\phi(m) = \frac{1}{(2\pi)^{1/2}\sigma} e^{-(m-m_0)^2/2\sigma^2} \tag{1.43}$$

where $\phi(m)$ reaches a peak at the *turnover point* m_0 and has a *dispersion* σ. (*NB:* Strictly defined, ϕ is the *relative* number of clusters – the number per unit magnitude, as a fraction of the total cluster population over all magnitudes – so that $\int \phi(m)dm = 1$. Often, however, it is plotted just as number of clusters per magnitude bin, in which case the proportionality constant in front of the Gaussian exponential must be renormalized.)

It must be stressed that this "Gaussian paradigm" has no physical basis other than the rough evolutionary scenarios sketched out above. It was adopted – rather informally at first – as a fitting function for the Milky Way and M31 globular cluster systems during the 1970's by Racine, Hanes, Harris, de Vaucouleurs and colleagues; it was finally established as the preferred analytical fitting function in two papers by Hanes (1977) and de Vaucouleurs (1977). Later, it became clear that the globular clusters in other galaxies consistently showed the same basic GCLF features that we have already discussed for the Milky Way, and the Gaussian description continued to be reinforced. There is, however, no astrophysical model behind it, and it remains strictly a descriptive function that has only the advantages of simplicity and familiarity.

Another analytic curve that is just as simple as the Gaussian is the t_5 function (Secker 1992),

$$\phi(m) = \frac{8}{3\sqrt{5}\pi\sigma} \left(1 + \frac{(m-m_0)^2}{5\sigma^2}\right)^{-3} . \tag{1.44}$$

Like the Gaussian, t_5 is a symmetric function with two free fitting parameters (m_0, σ) but differs slightly from the Gaussian primarily in the wings of the distribution. Secker's objective tests with the Milky Way and M31 GCLFs showed that t_5 is slightly superior to the Gaussian in providing a close match to the data. Applications to GCLFs in giant elliptical galaxies have also shown the same thing (Secker & Harris 1993; Forbes 1996a,b; Kissler et al. 1994).

The Gaussian model, however, has the strong advantage of familiarity to most readers.

More careful inspection of the full GCLF reveals, of course, that it is *not* symmetric about the turnover point – the long tail on the faint end of the distribution has no equivalent on the bright end, and cannot be matched by a single Gaussian or t_5 curve. To take account of these features, other authors have tried more complex polynomial expansions. Abraham & van den Bergh (1995) introduced

$$h_j = const \sum_{i=1}^{N} e^{-x_i^2/2} H_j(x_i) \tag{1.45}$$

where H_j is the jth-order Hermite polynomial, $x_i = (m_i - m_0)/\sigma$, and σ is the Gaussian dispersion. Here (h_3, h_4) are significantly nonzero for (e.g.) the Milky Way GCLF and represent the skewness and nonGaussian shape terms. Still another approach is used by Baum et al. (1995, 1997), as an attempt to reproduce the combined Milky Way and M31 GCLFs. They adopt an asymmetric hyperbolic function

$$\log \frac{n}{n_0} \;=\; -(a/b)\,\left(b^2 + (m-A)^2\right)^{1/2} \;+\; g(m-A) \tag{1.46}$$

where A is nearly equal to the turnover point m_0 for mild asymmetry, and a, b, g are parameters to be determined by the (different) bright-end and faint-end shapes of the GCLF.

It is trivially true that analytic curves with more free parameters will produce more accurate fits to the data. But none of these functions leads to any immediate insight about the astrophysical processes governing the cluster luminosities and masses, and we will not pursue them further.

1.5.5 The LDF: Power-Law Forms

In all of the fitting functions mentioned above, the GCLF is plotted as the number of clusters per unit *magnitude*. This way of graphing the function (dating back at least to Hubble's time) turned out to be a historically unfortunate decision. A more physically oriented choice would have been to plot the number per unit *luminosity* (or equivalently, the number per unit mass once we multiply by the mass-to-light ratio). We will call this latter form the *luminosity distribution function* or LDF. If the masses of globular clusters are governed by a fairly simple process of formation, then we could reasonably expect on physical grounds that the LDF might look like a simple *power law* in number per unit mass. Power-law distributions of mass or size, of the differential form $dN/dM \sim M^{-\alpha}$, are produced by many phenomena such as turbulence spectra, accretional growth, impact cratering, and so forth; and we will see strong evidence later in our discussion that the same result is true for star clusters that have recently formed in active galaxies.

Early suggestions that the LDF of globular clusters did in fact obey a power-law form were made by Surdin (1979) and Racine (1980) but these were unfortunately ignored for many years. The point was rediscovered by Richtler (1992). The power-law formulation of the LDF was pursued further by Harris & Pudritz (1994) and connected for the first time to a specific physical model of cluster formation in giant gas clouds (resembling Searle-Zinn fragments; we will return to a discussion of formation modelling in Sections 8-9).

How does the LDF relate to our more familiar GCLF? Considerable confusion between them persists in the literature, even though the difference between them is mathematically trivial. The GCLF histogram uses bins in magnitude (log L), whereas the LDF uses bins in L itself. Obviously, equal intervals in log L are not equal intervals in L, so the number distribution per bin has a different shape in each case. The power-law LDF is essentially a plot of $N(L)dL$, whereas the Gaussian-like GCLF $\phi(M_V)$ is a plot of $N(\log L)d(\log L)$. Thus the two functions scale as

$$\phi(M_V) \sim \frac{dN}{d\log L} \sim L\,\frac{dN}{dL} \sim L^{1-\alpha}\,.$$

The relation between the two forms is discussed in detail by McLaughlin (1994).

A power-law fit to the Milky Way LDF is shown in Fig. 1.37, adapted from Harris & Pudritz (1994). An exponent $\alpha_1 = 1.8 \pm 0.2$ provides an entirely acceptable fit for almost the entire observed range of cluster masses. It is only over the lowest $\sim$ 10% (i.e. $M \lesssim 10^5 M_{\odot}$) that the curve predicts far more clusters than are actually present. For this lower mass range, the data follow a much shallower power law near $\alpha_0 \sim 0.2 \pm 0.2$, at least in the Milky Way and M31 (Harris & Pudritz 1994; McLaughlin 1994). It is precisely at this rather abrupt slope change between α_0 and α_1 at $\sim 10^5 M_{\odot}$ that the classic turnover point of the GCLF lies. About half the numbers of clusters are fainter than that point, and about half are brighter. But it is empirically true that the GCLF $\phi(M_V)$ is roughly *symmetric* about the turnover point. In that case, we then require the exponents α_0, α_1 to be related by $(1-\alpha_0) \simeq (\alpha_1 - 1) \simeq 0.8 \pm 0.2$.

As noted earlier, we should expect that the LDF (or GCLF) as we see it today is a relic of *both* the mass spectrum of cluster formation and the subsequent dynamical evolution of the system. If we look at the distinctive and simple shape of the LDF – a double power law with a fairly sharp transition at $\sim 10^5 M_{\odot}$ – it is natural and extremely tempting to speculate that the *upper* mass range (α_1) represents the mass distribution laid down at formation, while the *lower* range (α_0) is created by the long-term effects of dynamical evolution, which should gradually carve away the more vulnerable lower-mass clusters. However, there are hints from observations of young globular clusters in interacting and merging galaxies that the initial mass distribution *already* has a shallow slope at the low-mass end right from the start; see Section 9 below. We are still some crucial steps away from understanding the correct full set of initial conditions for the LDF.

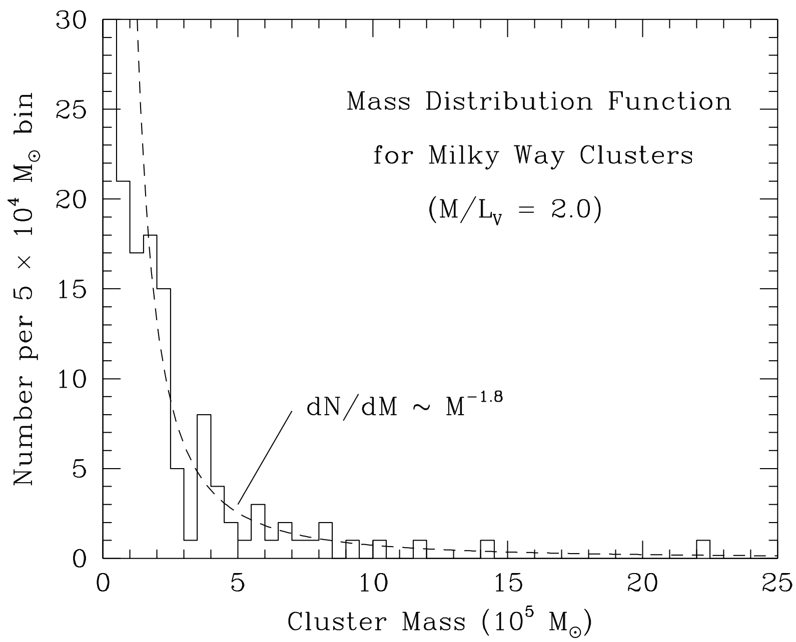

Fig. 1.37. Number of globular clusters per unit mass, plotted in linear form. The directly observed cluster luminosity L_V in solar units, given by $L_V = 10^{0.4(M_{V\odot} - M_V^T)}$ has been converted to mass $M/M_\odot$ with constant mass-to-light ratio $M/L_V = 2.0$. A power law $dN/dM \sim M^{-1.8\pm0.2}$ matches the data for $M \gtrsim 2 \times 10^5 M_\odot$ but continues upward too steeply for lower masses

If this rough scenario of formation combined with evolution is indeed on the right track, then we should expect that large numbers of low-mass clusters have been destroyed over the past Hubble time. But does that mean that a large fraction of the Galactic halo was built from dissolved globular clusters? No! The important point here is that these small clusters contain a *small fraction of the total mass* in the whole cluster system, even if they are numerous. The big clusters simply outweigh them by large factors. (For example, ω Centauri, the most luminous cluster in the Milky Way, contains about 8% of the total GCS mass all by itself. The clusters more massive than $10^5 M_\odot$ contain almost 95% of the total.) Harris & Pudritz (1994) demonstrate that if the original mass distribution followed the α_1 slope all the way down to low mass, then only about 30% of the total mass of the GCS would be lost if *every* cluster less massive than the present-day turnover point were destroyed.

Accounting for the fact that the clusters at all masses will be at least partially damaged by dynamical erosion (cf. the references cited above), it is reasonable to guess that roughly half the total original mass in the GCS has been escaped from the clusters to join the field halo population (see also Gnedin & Ostriker 1997; Murali & Weinberg 1997c; Vesperini 1998, for similar mass ratio estimates from dynamical simulations). The field halo stars outnumber the cluster stars by roughly 100:1, so most of the field stars must have originated in star-forming regions that never took the form of bound star clusters.

1.5.6 Comparisons with M31

A crucial step in constraining formation models for globular clusters in the early protogalaxies – as well as dealing with the more practical matter of using them as standard candles – is to start comparing the Milky Way GCS with those in other galaxies. The first and most obvious test is with M31, as the nearest large galaxy reasonably similar to ours.

M31 contains a globular cluster population at least twice as large as the Milky Way's, and thus more than all the other Local Group galaxies combined. The first ~ 150 were found by Hubble (1932) in a photographic survey. Major additions to the list of globular cluster candidates were made in subsequent photographic surveys by Baade (published in Seyfert & Nassau 1945), Vetesnik (1962), Sargent et al. (1977), Crampton et al. (1985), and by the Bologna consortium in the early 1980's, leading to their complete catalog (Battistini et al. 1987) which lists ~ 700 candidates of various quality rankings. About 250 – 300 of these are almost certainly globular clusters, while another ~ 300 are almost certainly various contaminants; the remainder (mostly projected on the M31 disk) have not yet been adequately classified. The true GCS population total in M31 is thus probably around 300, but is simply not yet known to better than $\sim 30\%$.

The published surveys used a wide range of strategies including object image morphology, brightness and color, and low-resolution spectral characteristics. The object lists from the different surveys overlap considerably in their discoveries and rediscoveries of clusters, but also contain numerous contaminants. Projected on the disk of M31, there are also open clusters, compact HII regions, and random clumps of bright stars which can masquerade as globular clusters. In the vast halo beyond the disk area, the main interlopers are faint, distant background galaxies which may have the round and compact morphology of globular clusters. The published surveys cover the M31 region rather thoroughly out to projected distances of only $r \sim 20$ kpc; a few more remote clusters have been found (Mayall & Eggen 1953; Kron & Mayall 1960), but the outer halo – a huge projected area on the sky – has yet to be systematically searched. If the Milky Way situation is any indication, not many new remote ones can be expected. A new CCD imaging survey primarily covering the disk of M31 by Geisler et al. is underway, and promises to push the detection limits to the faint end of the GCLF.

The true globulars in M31 must be found one by one. The techniques that have proven most useful include:

- *Radial velocity* (Huchra et al. 1991): Since $v_r(\mathrm{M31}) = -350$ km s^{-1}, any objects with strongly positive v_r are certainly background galaxies or (occasionally) Milky Way halo stars.
- *Image structure* (Racine 1991; Racine & Harris 1992): high resolution imaging with a large telescope can produce a definitive classification: if a candidate object is resolved into stars, it is a cluster. Under $\sim 0\farcs5$ resolution with the CFHT, color-magnitude diagrams for several halo clusters were successfully achieved which showed directly for the first time that the M31 clusters were indeed "normal" globulars like the familiar ones in the Milky Way (Heasley et al. 1988; Christian et al. 1991; Couture et al. 1995). With the greater resolving power of the HST, deeper and more precise CMDs for the M31 clusters have now become possible (Rich et al. 1996; Fusi Pecci et al. 1996; Holland et al. 1997).
- *Color indices*: The integrated colors of globular clusters occupy a fairly narrow region in a two-color plane such as $(U-B, B-V)$, $(B-V, V-R)$, etc. Other types of objects – galaxies particularly – usually have very different colors and can be eliminated. Reed et al. (1992, 1994) show that two-thirds of the background galaxies can be cleanly rejected this way.

Using a combination of all three of these techniques, Reed et al. (1994) constructed an almost completely clean sample of halo clusters (not projected on the M31 disk, thus free of internal reddening differences) which provides the best database we have for comparison with the Milky Way GCS.

In most other respects – metallicity distribution and kinematics – the M31 globular cluster system presents much the same story as does the Milky Way system, with small differences. Huchra et al. (1991) define MRC and MPC populations separated at [Fe/H] $= -0.8$ which strongly resemble the analogous ones in the Milky Way. The MRC is the more centrally concentrated of the two, and has a healthy overall rotation reaching ~ 150 km s^{-1} at a projected radius in the disk $r \simeq 5$ kpc. The MPC is more spatially extended and has much smaller net rotation ($\lesssim 50$ km s^{-1}). It is, however, also true that the overall scale size of the M31 system is bigger than that of the Milky Way, and moderately metal-rich clusters can be found at surprisingly large distances (the luminous cluster G1 = Mayall II is the prime example, with [Fe/H] ~ -1 at a projected distance of 40 kpc). The *relative* numbers of clusters at different metallicities, at least in the halo, are not strongly different from the [Fe/H] distribution in the Milky Way (see Fig. 1.38). Huchra et al. (1991) find relatively more metal-rich objects overall, including those in the disk, but it is easy to recognize the same bimodal form of the metallicity distribution, broadened by observational scatter.

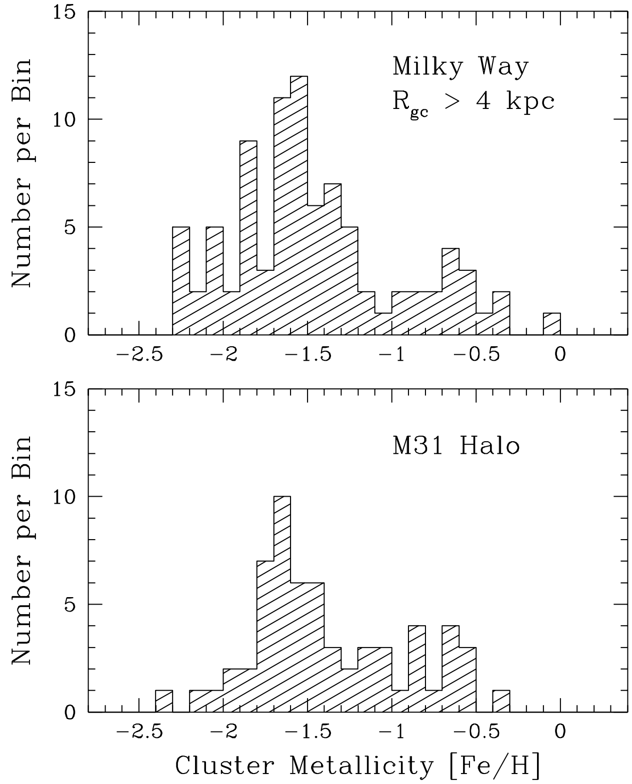

Fig. 1.38. Metallicity distribution for globular clusters in two galaxies. (**a**) In the upper panel, the distribution is shown for the Milky Way clusters with $R_{gc} > 4$ kpc. (**b**) In the lower panel, the distribution for the M31 halo is shown, with data drawn from Reed et al. (1994). The M31 sample is only slightly more weighted to MRC objects

There is still much work to be done to understand the characteristics of the M31 GCS at the same level of detail as we have for the Milky Way, and readers should refer to the papers cited above for further discussion.

The first GCLF comparison between M31 and the Milky Way was done by Hubble himself, in his 1932 discovery paper. At that time, the standard distance modulus in use for M31 was $(m - M) = 22$, almost three magnitudes smaller than today's best estimates. To make matters worse, the adopted luminosity for the RR Lyrae stars then was $M_{pg} \simeq M_B = 0.0$, almost a magnitude brighter than today's calibrations. In other words, the M31 clusters were being measured as much too faint in absolute magnitude, the Milky Way clusters too bright, and the combination left almost no overlap between the two GCLFs. Hubble relied on the similar *form* of the distributions as much as on their luminosity levels, and successfully concluded that "among known types of celestial bodies, the objects in M31 find their closest analogy in globular clusters".

The next serious GCLF comparison was in the landmark paper of Kron & Mayall (1960), which presented a comprehensive new set of integrated magnitudes and colors for globular clusters in several Local Group galaxies. By then, the basic distance scale issues in the Local Group had been settled (at least, the M31 discrepancy had been reduced to ~ 0.5 mag, rather than the 2.5-mag difference used during Hubble's time), and the true similarity between the globular cluster systems in M31 and the Milky Way clearly emerged. It was, by that time, also evident that the peak point M_V^0 of the GCLF was not just an artifact of incomplete observations, but was a real feature of the GCSs. Kron & Mayall's paper represents the *first explicit use of the GCLF turnover point as a standard candle for distance determination*. In all respects it is the same approach as we use today (Jacoby et al. 1992).

What does the M31/Milky Way comparison look like in modern terms? Unfortunately, we have to restrict our match to the halos of each, since the available list of objects projected on the disk of M31 is still too ill-defined (it is too contaminated with non-globulars, too incomplete at faint magnitudes, and photometrically too afflicted with random errors and poorly determined differential absorption). However, the halo sample (Reed et al. 1994) gives us an excellent basis for comparison: it is clean, complete down to a magnitude level well past the turnover point, unaffected by differential reddening, and well measured by modern CCD photometry.

We first need to worry a bit more, though, about which part of the Milky Way system we should use for comparison. We have already seen (Fig. 1.36 and Table 1.9) that its GCLF parameters depend on location. Another way to display it, using the running mean approach defined in our kinematics discussion, is shown in Fig. 1.39. The average luminosity (and also the turnover point, which is the GCLF median) rises smoothly outward from the Galactic center to a maximum at $R_{gc} \simeq 9$ kpc, then declines again. How can we make a valid comparison in the face of this amount of internal variation, which seems to vitiate the whole use of the GCLF as a standard candle?

But wait! The Milky Way is unique in the sense that it is the only galaxy for which we have the full three-dimensional information on the GCS space distribution. As a indicator of dynamical effects on the system or even small differences in the typical cluster mass at formation, Fig. 1.39 is of great interest on its own merits. But it should not be applied as it stands to any other galaxy. Instead, we should look at the Milky Way as if we were far outside it and could see only the distances of the clusters projected on the sky. The best projection to use is $r_p = \sqrt{Y^2 + Z^2}$, dropping the $X-$axis which is most affected by internal distance errors (see Sect. 1 above). When we do this, and again take running means of cluster luminosity, we get the result shown in Fig. 1.40. Rather surprisingly, we see that the large-scale global variation has largely been smoothed out, and the biggest part of it (the first outward rise) has been compressed down to the innermost ~ 2 kpc projected onto the Galactic bulge.

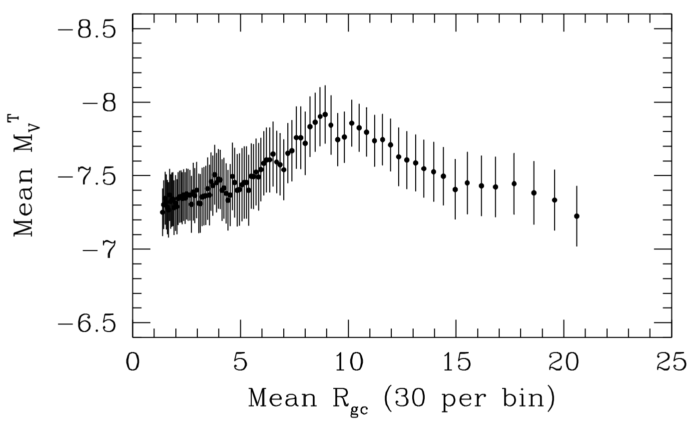

Fig. 1.39. Mean absolute magnitude for Milky Way globular clusters as a function of Galactocentric distance. Each point represents the mean $\langle M_V^T \rangle$ for the 30 clusters centered at the given distance; the next point outward is the same mean where the innermost cluster in the previous bin has been dropped and the next cluster outward has been added

Somewhat arbitrarily, I will take the region $r_p > 3$ kpc (containing 75 clusters) as the fiducial Milky Way sample. If we were to view the Milky Way at the same inclination angle to the disk as we see M31, this cutoff in projected distance would correspond roughly to the inner distance limits in the M31 halo sample. Over this range, there is little variation in the Milky Way mean cluster luminosity, and we can be more encouraged to try it out as a standard candle. We have just seen that this uniformity is something of an illusion! Larger internal differences are being masked, or washed out, by the projection effect from three to two dimensions. But the Galaxy can hardly be unique in this respect. We must therefore suspect that *the same smoothing may well be happening for the GCLF in any large galaxy that we look at*, where much of the information on the true amount of internal variation with position has simply been lost. To my knowledge, this point has not been realized, or used, in any previous application of the Milky Way GCLF as a distance indicator.

Having set up the fairest comparison sample that we can manufacture from the Milky Way, we can finally match it up with M31. The result is shown in Fig. 1.41. These two GCLFs are remarkably similar. The M31 sample

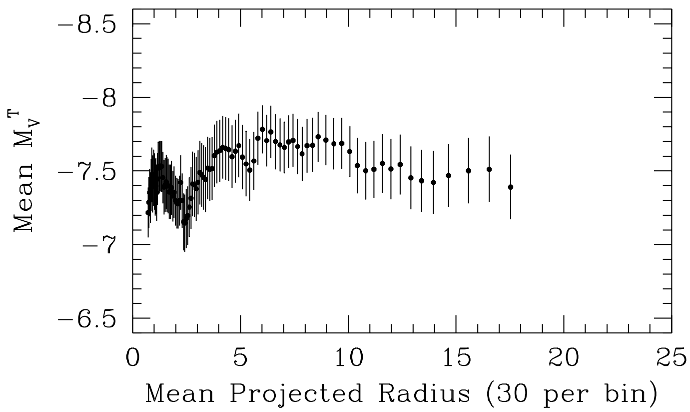

Fig. 1.40. Mean absolute magnitude for Milky Way globular clusters as a function of two-dimensional projected distance from the Galactic center, $r_p = (Y^2 + Z^2)^{1/2}$. Each point represents the mean $\langle M_V^T \rangle$ for the 30 clusters centered at the given distance; the next point outward is the same mean where the innermost cluster in the previous bin has been dropped and the next cluster outward has been added

differs only in the lack of faint clusters ($M_V^T \gtrsim -5.5$), for which the existing surveys are incomplete. Fitting Gaussian or t_5 functions to both galaxies yields turnover levels of $M_V^0 = -7.68 \pm 0.14$ (for the Milky Way projected-halo sample) and $M_V^0 = -7.80 \pm 0.12$ (for the M31 halo sample). These numbers are not significantly different. Let us turn the argument around: if we had used the M31 GCLF to *derive* a distance modulus, we would have obtained $(m-M)_V = V_0(\mathrm{M31}) - M_V^0(\text{Milky Way}) = (17.00 \pm 0.12) - (-7.68 \pm 0.14) = (24.68 \pm 0.18)$. Thus at the $\sim 0.10 - 0.15$ magnitude level of precision, the GCLF turnovers are similar and the method seems to work much better than we had any right to expect.

In Section 7, we will start testing this procedure for much more remote galaxies, and eventually become bold enough to derive the Hubble constant with it. Next, though, we need to take our first steps beyond the Local Group and find out what globular cluster systems look like in galaxies of very different types.

1.6 AN OVERVIEW OF OTHER GALAXIES: BASIC PARAMETERS

First gather the facts; then you can distort them at your leisure.

Mark Twain

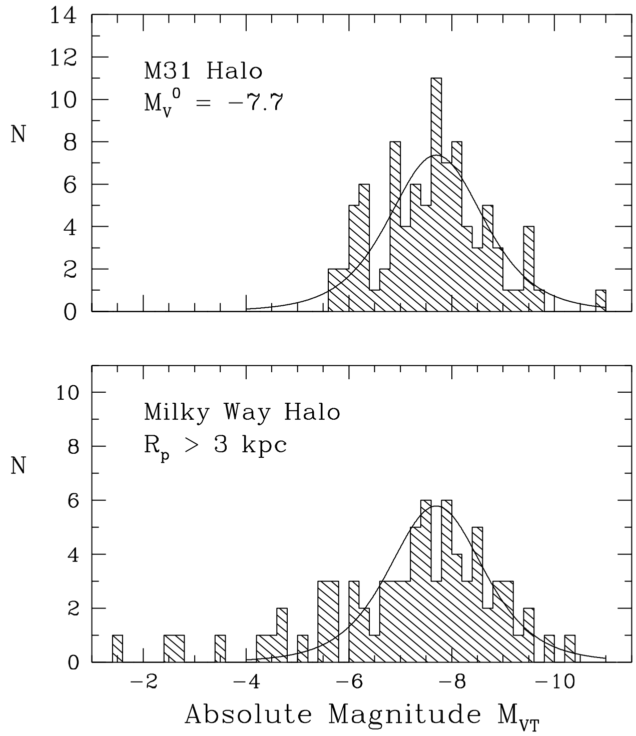

Fig. 1.41. Comparison of the GCLFs in M31 (upper panel) and the Milky Way (lower panel). The Milky Way sample is defined from clusters with projected (2-D) distances larger than 3 kpc as discussed in the text. The M31 halo sample, from Reed et al. (1994), has been shifted to absolute magnitude assuming our previously derived distance modulus $(m - M)_V = 24.80$ (Sect. 2)

Globular cluster systems have now been discovered and studied to some degree in more than a hundred galaxies. These cover the entire range of Hubble types from irregulars to ellipticals, the luminosity range from tiny dwarf ellipticals up to supergiant cD's, and environments from isolated "field" galaxies to the richest Abell clusters.

With the best imaging tools we have at present (the *HST* cameras), individual globulars can be resolved into their component stars rather easily for galaxies within the Local Group, i.e. at distances $\lesssim 1$ Mpc. With increasing difficulty, resolution of clusters into stars can also be done for galaxies up to several Megaparsecs distant (a color-magnitude diagram has been obtained for a halo cluster in the giant elliptical NGC 5128 at $d \simeq 4$ Mpc; see G.Harris et al. 1998). But for still more remote galaxies, we see the presence of the globular cluster population only as an excess of faint, small objects concentrated around the galaxy center (Fig. 1.42).

Studying the GCSs in most galaxies then becomes more of a statistical business, with the genuine clusters seen against a background of "contaminating" field objects (usually a combination of foreground stars and faint, compact background galaxies). Identifying *individual* globulars one by one can be done – usually radial velocity measurement is a definitive separator when combined with the integrated magnitudes and colors – but only with observational efforts that are greatly more time consuming.

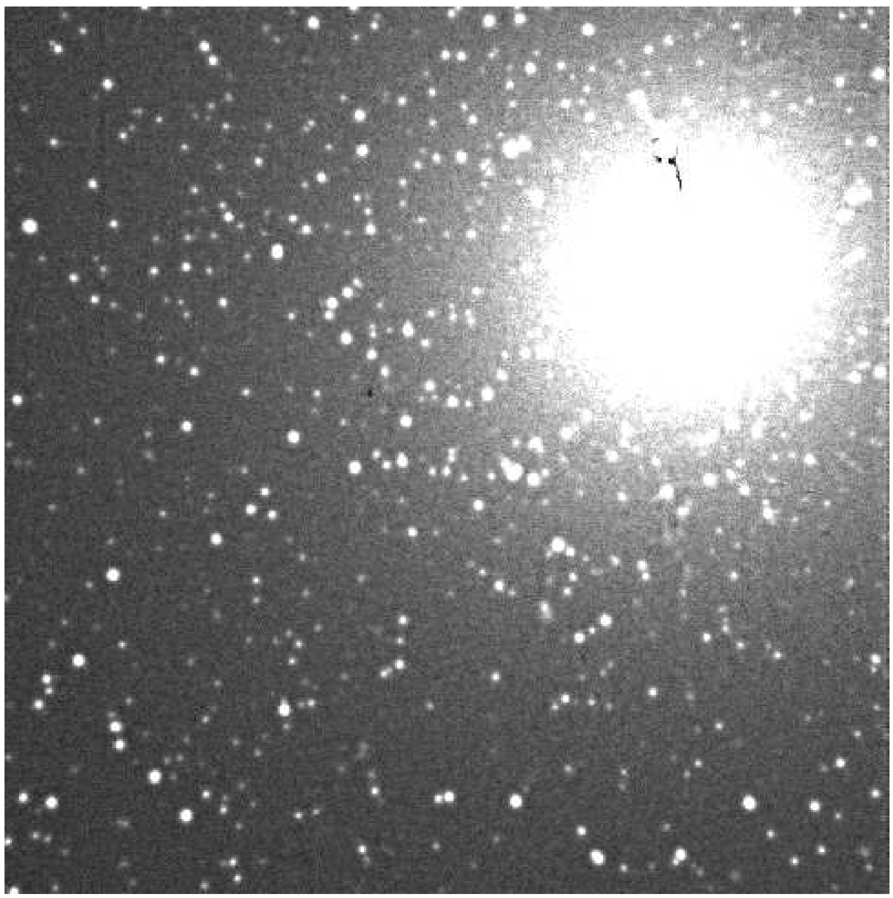

Fig. 1.42. A deep $R-$band image of the Virgo giant elliptical M87, with its swarm of globular clusters shown as the hundreds of faint starlike images around it. Traces of the nuclear jet can be seen extending upward from the galaxy center. The region shown is about 10 kpc on a side. This image was acquired with the High Resolution Camera at the Canada-France-Hawaii Telescope (see Harris et al. 1998a)

What characteristics of a GCS can we measure? The list of quantities, given below, is almost the same as for the Milky Way. Our only serious restriction is that the depth and quality of the information we can gather inevitably becomes more limited at larger distances.

Total populations: Simplest of all measurable quantities is the number of globular clusters present in the galaxy. Naively, we might expect that N_{cl} should rise in direct proportion to the galaxy luminosity (or mass). That seemed to be the case in the early days of the subject (Hanes 1977; Harris & Racine 1979), with the single exception of M87, which was recognized from the beginning to be anomalous, However, it is now realized that there is a great deal of real scatter from one galaxy to another around this basic proportionality relation. Understanding the total population size or global cluster formation efficiency is one of the most challenging questions in the entire subject, leading us far into issues of galaxy formation and evolution.

Metallicities: The normal assumption or "null hypothesis" of GCS work is that the old-halo clusters we see in other galaxies basically resemble the familiar ones in the Milky Way. This assumption has been fully borne out in the various Local Group members where detailed comparisons of stellar content have been possible. We can then use the integrated cluster colors in some reasonably sensitive index like $(V-I)$, $(C-T_1)$, etc., to estimate the *metallicity distribution function* (MDF) of the system, since metallicity determines the integrated color of old stellar systems much more strongly than other factors such as age. With quite a lot more effort, we can use the absorption line indices in their integrated spectra to do the same thing. Where it has been possible to do both, the color and spectral approaches for estimating metallicity agree well both at low dispersion (e.g., Racine et al. 1978; Huchra et al. 1991; Brodie & Huchra 1991) and in the finer detail that has been achieved more recently (e.g., Cohen et al. 1998; Jablonka et al. 1992, 1996).

Luminosities: The *luminosity distribution function* (LDF; see Section 5) of the GCS is the visible signature of the cluster mass distribution. As discussed earlier, the LDF we see today should be the *combined* result of the mass spectrum at formation, and the subsequent effects of dynamical evolution in the galactic tidal field.

Spatial distribution: The GCS in any galaxy is a centrally concentrated subsystem, generally following the structure of the visible halo light. However, particularly in giant ellipticals the GCS often traces a somewhat shallower radial falloff than the halo, and in extreme cases (cD galaxies) it may be closer to representing the more extended dark-matter potential well. Many recent studies have attempted to correlate the MDF and LDF with the spatial distribution, thus extracting more clues to the system formation and evolution.

Radial velocity distribution: In principle, much the same types of kinematic and dynamical studies of the Milky Way GCS can be carried out in any galaxy for which we can acquire a large enough set of cluster radial velocities. However, the internal precisions of the velocity measurements need to be $\sim \pm 50$ km s^{-1} for the internal dynamics of the halo to be adequately studied, and acquiring absorption-line velocities for large samples of objects as faint as those in Virgo and beyond has been difficult. With the advent of the new generation of 8-m and 10-m optical telescopes, this type of work has now been started in earnest by several groups.

1.6.1 Defining and Measuring Specific Frequency

The total population of clusters in a galaxy is usually represented by the *specific frequency* S_N, the number of clusters per unit galaxy luminosity (Harris & van den Bergh 1981; Harris 1991):

$$S_N = N_{cl} \cdot 10^{0.4\,(M_V^T + 15)} \tag{1.47}$$

where M_V^T is the integrated absolute magnitude of the host galaxy and N_{cl} is the total number of clusters. This definition can be rewritten in terms of the visual luminosity of the galaxy L_V in solar units,

$$S_N = 8.55 \times 10^7 \, \frac{N_{cl}}{(L_V/L_\odot)} \, . \tag{1.48}$$

Estimating S_N for a given galaxy is therefore a simple process in principle, but requires two kinds of completeness corrections. If the imaging coverage of the galaxy is spatially incomplete, then radial extrapolations have to be made to estimate N_{cl}. Similarly, if the photometric limit reaches only part way down the GCLF (as is almost invariably the case), then an extrapolation in magnitude is also needed, starting with an assumed distance to the galaxy (Fig. 1.43). *By convention*, the GCLF shape is assumed to be Gaussian (see the previous section) for purposes of estimating the total population. In most galaxies, it is often the case that the faint limits of the observations turn out to be somewhere near the GCLF turnover.

There are two obvious ways in which the predicted value of N_{cl} can go wrong. If the fainter half of the GCLF has a very different shape from the brighter half that is directly observed, then we would end up miscalculating the total N_{cl}. And if the limit of observations falls well short of even the turnover point, then the extrapolation from $N(obs)$ to N_{cl} can be uncomfortably large even if the assumption of symmetry is valid. Thus it seems that the specific frequency is a rather uncertain number.

Or is it? The procedure is actually not as risky as it first looks, for two reasons:

- We calculate N_{cl} essentially by using the Gaussian-like shape of the GCLF to determine the number of clusters *on the bright half*, and then doubling it. In most galaxies we never see the faint half, and never use it. In other words, the specific frequency is really a ratio which compares the number of *bright* clusters in different galaxies. Thus the first rule of specific frequency is:

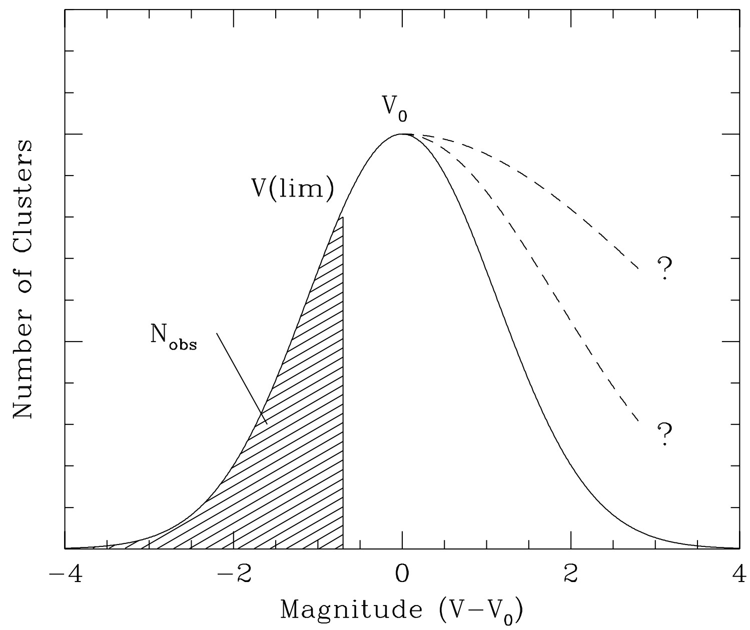

Fig. 1.43. Calculation of total cluster population. The GCLF is assumed to have a Gaussian-like form shown by the solid curve, with turnover point at apparent magnitude V_0. The limiting magnitude of the photometry is $V(lim)$, so that the total observed population of clusters N_{obs} is given by the shaded area. The total population N_{cl} over the entire GCLF is then $N_{cl} = N(obs)/F$ where the completeness fraction F is the shaded area divided by the total area under the Gaussian. If the unobserved faint half of the GCLF had a different shape (dashed lines), the total population estimate would be affected significantly, but the number of *bright* clusters (more luminous than the turnover) would not

S_N measures the number of clusters brighter than the GCLF turnover V_0.

- Despite this reassurance, S_N would still be an invalid quantity if the *absolute* magnitude of the turnover differed wildly from one galaxy to the next – or, indeed, if there were no turnover at all. But by all available evidence (introduced in Section 5, and discussed further in Section 7 below), the GCLF has amazingly similar parameters from place to place. Perhaps against all *a priori* expectations, the GCLF shape is the closest thing to a universal phenomenon that we find in globular cluster systems. Thus we have our second rule,

S_N provides a valid basis for comparison among galaxies because of the universality of the GCLF.

In summary, we can go ahead and use S_N knowing that it has reasonable grounding in reality.

The estimated S_N is fairly insensitive to the assumed galaxy distance d, because any change in d will affect both the calculated galaxy luminosity and total cluster population in the same sense (see Harris & van den Bergh 1981). However, it is sensitive to mistakes in the assumed limiting magnitude V_{lim}, or in the background contamination level. Suppose that for a distant galaxy you count N faint starlike objects around the galaxy, and N_b background objects in an adjacent field of equal area down to the same limiting magnitude. By hypothesis, the excess $N_o = (N - N_b)$ is the globular cluster population, and the uncertainty is

$$N_o \pm \Delta_N = (N - N_b) \pm \sqrt{N + N_b} \,. \tag{1.49}$$

The total over all magnitudes is $N_{cl} = N_o/F$. Now suppose that the uncertainty ΔV in the limit V_{lim} translates into an uncertainty ΔF in the completeness factor F: we can then show

$$\frac{\Delta S_N}{S_N} = \left(\frac{(N + N_b)}{(N - N_b)^2} + \left(\frac{\Delta F}{F} \right)^2 \right)^{1/2} \,. \tag{1.50}$$

Numerical trials with this relation show that to produce S_N estimates that are no more uncertain than (say) 20%, we need to have observations reaching $V_{lim} \gtrsim (V_0 - 1)$, i.e. to within a magnitude of the turnover or fainter. If the raw counts are *very* dominated by background contamination, the situation may be worse. And if the observations fall short of the turnover by 2 magnitudes or more, the relative uncertainty $\Delta S_N/S_N$ starts increasing dramatically and the estimates become quite rough.

1.6.2 Specific Frequency: Trends and Anomalies

Let us now turn to some of the results for specific frequencies. Elliptical galaxies are the simplest to work with, and make up by far the biggest share of the database for globular cluster systems. Fig. 1.44 shows the current results for E galaxies over all luminosities, from dwarfs to supergiant cD's (data are taken from the compilations of Blakeslee et al. 1997; Harris et al. 1998a; Miller et al. 1998; and a few recent individual studies).

From this simple graph we can already draw several conclusions. First, over a range of almost 10^4 in galaxy luminosity L, the mean specific frequency is nearly constant; that is, to first order the total number of clusters rises in nearly direct proportion to parent galaxy luminosity, $N_t \sim L$.

Second, there is *significant scatter* at all L. For giant ellipticals, S_N in individual galaxies ranges from a high near ~ 15 to a low near ~ 1 or perhaps even lower. For dwarf ellipticals, the range is even larger, with S_N(max) near 30. This scatter extends far beyond the internal uncertainties in estimating S_N, and must certainly be real. It was suspected to exist from the earliest samples of E galaxies (Hanes 1977; Harris & van den Bergh 1981), and later studies from more comprehensive samples and deeper photometry

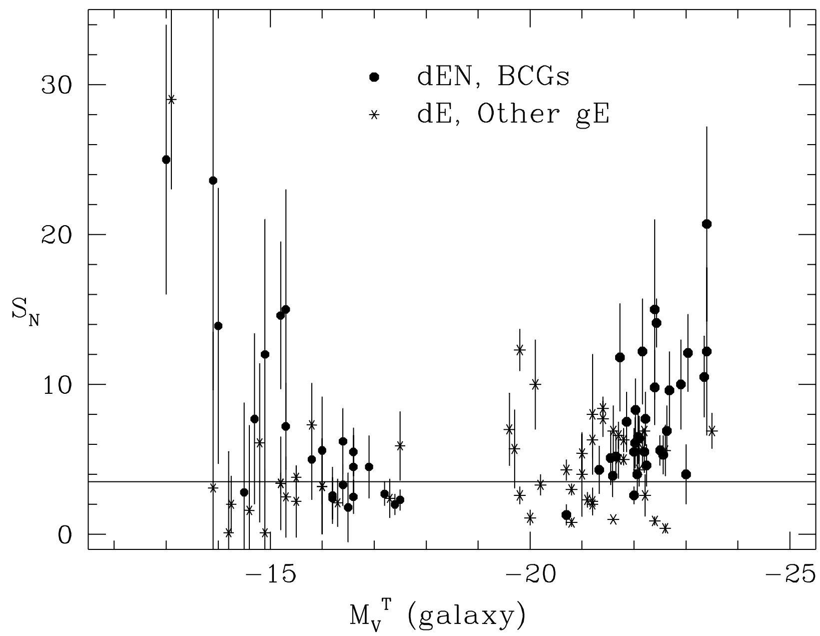

Fig. 1.44. Specific frequency S_N plotted against luminosity for elliptical galaxies. Solid symbols are for cD-type giants (brightest cluster ellipticals) and nucleated dwarf ellipticals, while starred symbols are for normal gE's and non-nucleated dwarfs. The baseline "normal" level is at $S_N = 3.5$; see text. The gap in the range $M_V^T \sim -18$ to -20 is a selection effect; no globular cluster systems have been studied for galaxies in that range

have only confirmed and extended the first estimates of the range in S_N that real galaxies exhibit. The specific frequency is a parameter which differs by as much as a factor of *twenty* between galaxies which have otherwise similar structures, luminosities, and metallicities. This is one of the most remarkable results to emerge from the study of globular cluster systems. Although some plausible ideas are beginning to emerge (Section 8), it still lacks a compelling theoretical explanation.

Third, significant correlations of S_N with other galaxy properties do exist. The strongest and most obvious connection is with *environment*. At the high-luminosity end, it appears that there is something special about the giant ellipticals that sit at the centers of large clusters of galaxies – the "brightest cluster galaxies" (BCGs) which often have cD-type structures (high luminosities, along with extended envelopes of stellar material that appear to follow the potential well of the cluster as a whole). These particular galaxies have the highest known specific frequencies among gE galaxies. The prototype of this class is M87, the Virgo cluster cD and the center of the biggest concentration of galaxies in the Virgo region, which has an extremely well determined

$S_N = 14.1 \pm 1.6$ (Harris et al. 1998a) almost three times larger than the mean for other Virgo ellipticals. Since cD's are few and far between, it took many years for a significant sample of globular cluster system observations to be built up for them, and for a long time M87 was regarded as being virtually unique (see, e.g., Hanes 1977; Harris & Smith 1976; Harris & van den Bergh 1981; Harris 1988a for the initial historical development). Since then, many more BCGs have been studied, and a clear correlation of S_N with luminosity has emerged: the more luminous BCGs have higher specific frequencies (Blakeslee 1997; Harris et al. 1998a). Since the brighter BCGs tend to be found in in more populous clusters of galaxies, the hint is that denser, richer environments lead to higher specific frequencies.

Even without the BCGs, a similar conclusion would emerge from the rest of the ellipticals. It was first suggested by Harris & van den Bergh (1981) that the ellipticals in small, sparse groups of galaxies or in the "field" had systematically lower S_N than those in richer systems like Virgo or Fornax. Larger samples confirmed this. For E's in small groups, S_N is typically $\sim 1{-}3$, while in larger groups (Fornax, Virgo, and above) we find $S_N \simeq 5$ (Harris 1991).

Until the past few years, not much was known about globular cluster systems in *dwarf* ellipticals, the small galaxies at the opposite end of the luminosity scale. But they, too, exhibit a large S_N range and some intriguing correlations which have been revealed by new surveys (Durrell et al. 1996a,b; Miller et al. 1998). There appears to be a dichotomy between *nucleated* dE's (those with distinct central compact nuclei) and non-nucleated dE's. The dE's present a fairly simple story, with a mean $\langle S_N \rangle \simeq 2$ independent of luminosity and with not much scatter. In striking contrast, the dE,N systems show a clear correlation of S_N with luminosity, in the opposite sense to the BCGs: less luminous dE,N's have higher specific frequencies. The most luminous dwarfs of both types have similarly low specific frequencies, but at progressively lower L, the specific frequency in dE,N's steadily increases, reaching the highest values at the low-L end.

We would like to understand why the specific frequency displays such a large range. How can otherwise-similar galaxies make (or keep) vastly different numbers of old-halo star clusters? Speculations began as soon as the phenomenon was discovered, concentrating first on the environmental connection and on the "anomaly" of the BCGs (e.g., Harris & Smith 1976; van den Bergh 1977; Harris & Racine 1979). Many other ideas entered the game later on. We will discuss these in the last two Sections; but for the moment, we will say only that no single explanation or mechanism seems able to produce the full range of specific frequencies seen amongst all the ellipticals. It is a remarkably simple phenomenon, but remains a hard one to explain.

By contrast with the ellipticals, disk and spiral galaxies so far present a much more homogeneous picture. The Sb/Sc/Irr systems, to within factors of two, have specific frequencies similar to that of the Milky Way, in the range $S_N \simeq 0.3 - 1.0$ (Harris 1991; Kissler-Patig et al. 1999). (In fact, given the difficulty in measuring S_N in disk-type galaxies, where the total numbers of clusters are much lower than in gE galaxies to start with, and where the disk light and dust add further confusion, it is possible that the nominal differences in specific frequency among disk galaxies are entirely due to observational scatter.) Apparently, the spirals have not experienced the same range of formation processes or evolutionary histories that the ellipticals have.

If we take these numbers at face value, it would seem the spirals and irregulars are much less efficient at forming globular clusters than are most ellipticals. But an obvious difficulty in making the comparison is that these late-type galaxies have much higher proportions of "young" stellar populations which make them more luminous than ellipticals of the same mass. To correct for this effect, it has become customary to adjust the total luminosity of the galaxy to the "age-faded" value that it would have if all its stars evolved passively to $\gtrsim 10$ Gyr, like those of ellipticals (see Section 9 below). This correction must be done on an individual basis for each spiral or irregular, and adds a further uncertainty to the comparison. On average, this "renormalized" specific frequency falls in the range $S_N \sim 2 \pm 1$, which is similar to the typical values for dE (non-nucleated) galaxies, or large E galaxies in sparse groups, or S0 galaxies (which are disk systems free of dust or young stars). They still fall well short of the $S_N \sim 5$ level associated with gE members of rich groups, which in turn are lower than most BCGs.

These comparisons make it tempting to suggest that there is a "natural" level for S_N which is somewhere in the range $\sim 2-4$, applying to spirals, S0's, dwarf ellipticals, and many large ellipticals in a wide range of environments (refer again to Fig. 1.44, where a baseline $S_N = 3.5$ is shown). It would not be reasonable to expect this level to be *exactly* the same in all these galaxies, because the numbers of clusters are determined by formation efficiencies and dynamical evolution which are, at some level, stochastic processes. Some scatter is also introduced in the measurement process (see above), which in the worst cases can leave S_N uncertain by 50% or so. The major mystery has always been the extreme situations which go far beyond these normal cases: the BCGs, and the nucleated dwarfs.

1.6.3 Metallicity Distributions

An important trace of the early history of any galaxy is left behind in the metallicity distribution of its halo stars. Unfortunately, almost all galaxies are too remote for individual stars to be resolved, so what we know about their chemical composition is indirect, relying only on various averages over the MDF. This is where the GCS gives us a distinct advantage: the globular clusters are old-halo objects that can be found *one by one* in galaxies far too distant for any individual stars to be studied, including many unusual galaxy types. In these galaxies, we can derive a full *distribution function* of metallicity for the GCS, and not just the mean metallicity (Harris 1995).

Insofar as the GCS represents the halo field-star population, we can use this MDF to deduce the early chemical enrichment history of the system.

In Fig. 1.45, MDFs are shown for a representative sample of galaxies which cover the presently known range of mean metallicities. The metallicity range correlates strongly with galaxy size. In the dE's, almost all the clusters are low-metallicity objects, like the MPC clusters in the halo of our Milky Way with an average near [Fe/H] $\simeq -1.6$ (see Fig. 1.7). This observation fits in well with standard views of the early evolution of dwarf ellipticals, in which a small, isolated protogalactic gas cloud undergoes a single major burst of star formation, but ejects a large fraction of its gas in the process (e.g., Dekel & Silk 1986; Babul & Rees 1992). Since its tiny potential well cannot hold the gas ejected by the first round of stellar winds and supernovae, the heavy-element enrichment cannot proceed to completion and the "effective yield" of the enrichment is much lower than normal, leaving only metal-poor stars behind (Hartwick 1976). What is therefore more surprising is that these small ellipticals have any metal-rich clusters at all: two with [Fe/H] $\gtrsim -1$ are probably members of the Local Group dE's, and there are clear hints that others can be found, albeit in small numbers, with similar metallicities (see, e.g., Durrell et al. 1996a for the Virgo dwarfs). How did these few relatively metal-rich clusters arise in circumstances that are strongly biased against normal metal enrichment? The answers are not yet clear. They may simply represent rare instances where unusually dense pockets of the proto-dE got an early start and held its gas long enough for the local enrichment to proceed up to higher levels than normal. Alternately, they may represent a somewhat later and more minor epoch of star formation driven by late infall of gas or by the triggering of whatever residual gas was left in the system.

In somewhat larger and more complex galaxies – the Milky Way, M31, and normal ellipticals – a metal-poor component is usually present at roughly the same metallicity level as we find in the dwarfs, but a much more significant higher-metallicity population also appears and the MDF as a whole begins to look very broad. At the upper end of the scale, in some high-luminosity ellipticals such as NGC 3311 in the Hydra I cluster and IC 4051 in Coma, the MPC component almost disappears and we are left with only a metal-rich GCS (e.g., Secker et al. 1995; Woodworth & Harris 1999). The relative proportions of MPC and MRC components can differ quite noticeably from one host galaxy to another, even among otherwise similar galaxies, and in some ellipticals the MDF is narrow and not easily described as a mixture of metal-poor and metal-rich components (e.g., Ajhar et al. 1994; Kissler-Patig et al. 1997a).

The large differences in MDFs, coupled with the amazingly similar *luminosity* distribution functions of globular clusters in all galaxies, already put important constraints on formation models for globular clusters. Clearly, the GCS formation process must be a robust one which gives the same cluster mass spectrum *independent of the metallicity of the progenitor gas clouds.*

In addition, the MDFs already challenge our traditional, Milky-Way-bred notions that a "globular cluster" is prototypically a massive, old, *metal-poor* star cluster. It is not. By sheer weight of numbers and high specific frequency, a large fraction of all the globular clusters in the universe reside in giant ellipticals, and many of these are metal-rich, extending up to (and beyond) solar metallicity.

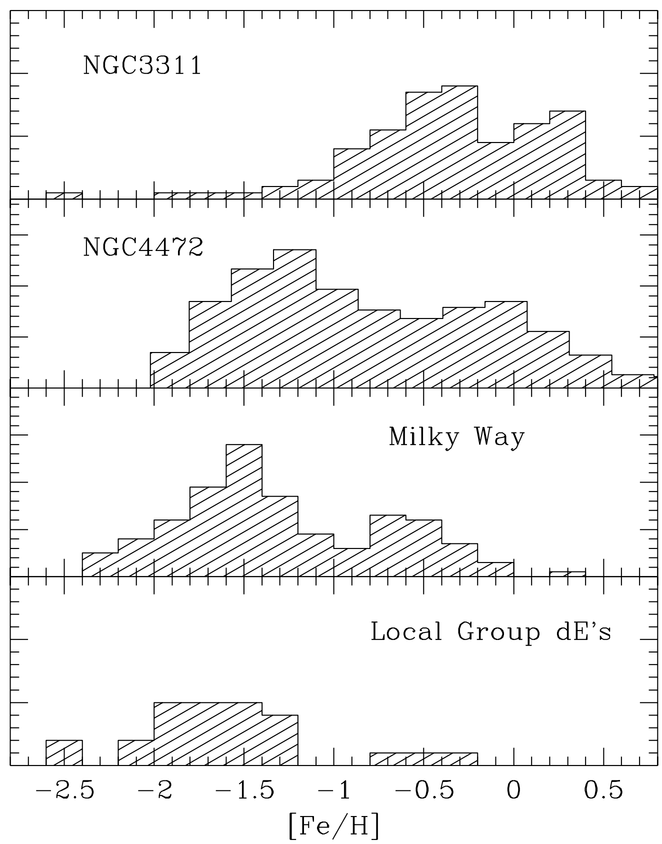

Fig. 1.45. Metallicity distribution functions for globular clusters in selected galaxies. The top panel shows the MDF for the cD galaxy in the Hydra I cluster (Secker et al. 1995); the second panel shows the Virgo giant elliptical NGC 4472 (Geisler et al. 1996); and the bottom panel is a composite of the clusters in all the Local Group dwarf ellipticals (Harris 1991; Da Costa & Armandroff 1995)

When the entire range of galaxies is plotted, we find a correlation of mean GCS metallicity with galaxy luminosity (Brodie & Huchra 1991; Harris 1991; Ashman & Zepf 1998; Forbes et al. 1996b). The equation

$$\langle[\mathrm{Fe/H}]\rangle = -0.17\, M_V^T - 4.3 \tag{1.51}$$

matches the overall trend accurately for the ellipticals. However, the correlation is much closer for the dE's than for the giant ellipticals, which exhibit a large galaxy-to-galaxy scatter in mean [Fe/H] and almost no trend with M_V^T. The reason for this large scatter appears (Forbes et al. 1997) to be that this mean correlation ignores the large variety of mixtures between the MPC and MRC parts of the MDF that are found from one galaxy to another. It seems too much of an oversimplification to think of the entire MDF as a unit.

Another general result valid for most E galaxies is that the GCS mean metallicity is *lower* than the galaxy halo itself by typically 0.5 dex. That is, the same scaling rule of metallicity versus total size applies to the galaxy itself and to the GCS, but with the GCS offset to lower metallicity (Brodie & Huchra 1991; Harris 1991). The initial interpretation of this offset (cf. the references cited) was that the GCS formed slightly earlier in sequence than most of the halo stars, and thus was not as chemically enriched. This view dates from a time when it was thought that there was a single, fairly sharply defined formation epoch for the clusters. As we will see below, however, the story cannot be quite that simple for most large galaxies.

1.6.4 Substructure: More Ideas About Galaxy Formation

We have already discussed the *bimodal* structure of the MDF for the Milky Way clusters: they fall into two rather distinct subgroups (MPC, MRC), and the MDF itself can be well matched analytically by a simple combination of two Gaussians. For giant E galaxies, it is easily possible to obtain MDFs built out of hundreds and even thousands of clusters, and the same sorts of statistical analyses can readily be applied. However, it was only during the past decade that MDFs for these galaxies became internally precise enough that bimodal, and even multimodal, substructure began to emerge from the obviously broad color distributions. Observationally, the most important breakthrough in this field was the employment of highly sensitive photometric indices, especially the Washington $(C - T_1)$ index (Geisler & Forte 1990). With it, the intrinsic metallicity-driven color differences between clusters stood out clearly above the measurement scatter for the first time, and CCD photometry of large samples of clusters could be obtained. (Other well known color indices such as $(B-V)$ or $(V-I)$ are only half as sensitive to metallicity as $(C-T_1)$ or $(B-I)$. Although it is still possible to obtain precise MDFs from them, the demands for high precision photometry are more stringent, and were generally beyond reach until the present decade; see Ashman & Zepf 1998).

On the analytical side, better statistical tools were brought to bear on the MDFs (Zepf & Ashman 1993; Ashman et al. 1994; Zepf et al. 1995). These studies revealed that the color distributions of the clusters in giant E galaxies, which were initially described simply as "broad", could be matched better as bimodal combinations of Gaussians strongly resembling the ones for the Milky Way. Improvements in the quality of the data have tended to confirm these conclusions, with the multimodal character of the color distribution standing out more clearly (e.g., Whitmore et al. 1995; Geisler et al. 1996; Forbes et al. 1998; Puzia et al. 1999). Since the integrated colors of globular clusters vary linearly with [Fe/H] for [Fe/H] $\lesssim -0.5$ (Couture et al. 1990; Geisler & Forte 1990), a bimodal color distribution translates directly into a bimodal MDF.

But is a bimodal MDF a clear signature of two major, distinct formation epochs in these gE galaxies, analogous to the ones postulated for the Milky Way? Zepf and Ashman have repeatedly interpreted the bimodality in terms of their merger model for elliptical galaxies (Ashman & Zepf 1992), in which the MPC clusters are assumed to be the ones formed in the first star formation burst, and the MRC ones are due to later bursts driven by mergers and accretions which bring in new supplies of gas. This scenario will be discussed further in Section 9 below. Meanwhile, other authors have noted that bimodality is, although common, not a universal phenomenon in E galaxies. In many other cases, the GCS color distribution is closer to a unimodal one and remarkably narrow, with typical width σ[Fe/H] $\simeq 0.3$ (e.g., Ajhar et al. 1994; Kissler-Patig et al. 1997a; Elson et al. 1998). Furthermore, in cases where the MDF is approximately unimodal, the *mean* metallicity of the clusters is not always the same between galaxies: some are rather metal-poor (like those in the Milky Way halo), while others (notably the Coma giant IC 4051 or the Hydra cD NGC 3311; see Secker et al. 1995; Woodworth & Harris 1999) are strikingly metal-rich, with a peak at [Fe/H] $\simeq -0.2$ and clusters extending well above solar abundance. In such galaxies, it is puzzling that there would be little or no trace of any first-generation metal-poor stellar population.

Forbes et al. (1997) provide an analysis of all the available MDFs for giant ellipticals which suggest an interesting pattern in the mean metallicities of the two modes. The peak of the MRC falls consistently at [Fe/H](MRC) $= -0.2$ for the most luminous gE's, with little scatter. By contrast, the peak of the MPC (on average ~ 1 dex lower) shows considerable galaxy-to-galaxy scatter. The cluster sample sizes in some of these galaxies are small, and the identifications of the mode locations are debatable in some cases; but their basic conclusion seems sound, and may turn into a strong constraint on more advanced formation models. They argue that a two-phase *in situ* burst is the best interpretation to generate the basic features of these MDFs.

The presence or absence of bi- or multi-modality seems to correlate with little else. Generally valid statements seem to be that the cD-type (BCG) galaxies have the broadest MDFs (bimodal or multimodal); normal ellipticals can have broad or narrow MDFs according to no pattern that has yet emerged. The sample of well determined MDFs is, however, not yet a large one, and could be considerably expanded with studies of more ellipticals in more environments and over a wider range of sizes.

A superb illustration of what can be obtained from such studies is found in the work by Geisler et al. (1996) and Lee et al. (1998) on the Virgo giant NGC 4472. With CCD imaging and Washington filters, they obtained accurate $(C-T_1)$ indices for a deep and wide-field sample of globular clusters around this gE galaxy. Two diagrams taken from their study are shown in Figs. 1.46 and 1.47. A plot of cluster color or metallicity against galactocentric radius (Fig. 1.46) reveals distinct MRC and MPC subpopulations which also follow different spatial distributions. The redder MRC objects follow a radial distribution that is similar to that of the halo light of the galaxy, and their mean color is strikingly similar to that of the halo (Fig. 1.47). By contrast, the bluer MPC clusters – equally numerous – follow a much more extended spatial structure and are more metal-poor than the MRC clusters or the galaxy halo by fully 1 dex.

The GCS *as a whole* displays a radial metallicity gradient, with the mean color decreasing steadily outward (Fig. 1.47). Yet neither the MPC or MRC subgroups exhibit significant changes in mean color with radius by themselves. The gradient in the GCS as a whole is, therefore, in some sense an artifact! It is a simple consequence of the different radial distributions of the two subpopulations: the MRC clusters dominate the total GC numbers at small radii, while the MPC clusters dominate at large radii, so that the mean color of all clusters combined experiences a net outward decrease.

This same observational material also allows us to place interesting limits on the *specific frequency for each of the two subgroups*. Taking the bimodal MDF at face value, let us suppose that NGC 4472 formed in two major starbursts. By hypothesis, the earliest one produced the MPC clusters along with some halo light (i.e., field stars) at the same metallicity. The later and stronger burst formed the MRC clusters, with more field stars and with associated clusters at the same (higher) metallicity.

Geisler et al. (1996) estimate that the total number of MPC clusters is $N_{MPC} = 3660$. Their mean color is $(C-T_1) = 1.35 \pm 0.05$ (see Fig. 1.47). Similarly, for the MRC clusters they estimate $N_{MRC} = 2440$, with a mean color $(C-T_1) = 1.85 \pm 0.05$. But now, the mean color of the *halo light* is $(C-T_1) = 1.85 \pm 0.05$, exactly the same as that of the MRC clusters. Thus under our assumptions, *the vast majority of the halo stars must belong to the second, more metal-rich formation epoch*; otherwise, their integrated color would lie distinctly between the two groups of clusters. A straightforward calculation shows that if the MPC halo light (which by hypothesis has a color $(C-T_1) \simeq 1.35$) makes up more than about 6% of the total galaxy light, then the integrated color of the whole halo will be bluer than $(C-T_1) = 1.80$, which would bring it outside the error bars of the observations.

Turning this calculation around, we conclude that the MRC starburst made up $\gtrsim 94\%$ of the stellar population of the galaxy.

Finally, we can convert these numbers into specific frequencies. The integrated magnitude of the whole galaxy is V^T(N4472) = 8.41. Splitting it in the proportions estimated above, we then have V^T(MRC) $\simeq$ 8.48, and V^T(MPC) $\gtrsim$ 11.30. Thus the metal-richer component has

$$S_N(MRC) = 2.4 \pm 0.3$$

while a *lower limit* for the metal-poor burst is

$$S_N(MPC) \gtrsim 50 \ !$$

The specific frequency of the first, metal-poor starburst must have been extremely high – higher, in fact, than in any galaxy as a whole that we know of today. Either the conversion rate of gas into bound globular clusters was outstandingly efficient in the initial burst, or a great deal of the initial gas present was ejected or unused for star formation during the burst. As we will see later, the latter explanation currently seems to be the more likely one (see also Forbes et al. 1997 for a similar argument). One possibility (Harris et al. 1998a) is that the initial metal-poor gas formed the MPC clusters that we now see, but was then prevented from forming its normal proportion of stars by the first major burst of supernovae and the development of a galactic wind. A large part of this gas – now enriched by the first starburst – later underwent dissipational collapse, most of it then being used up in the second burst. Contrarily, the specific frequency of the second starburst – which produced most of the galaxy's stars – was quite modest, falling well within the "normal" range mentioned previously for many kinds of galaxies.

This discussion operates within the context of a generic "in situ" model of formation, i.e., one in which the galaxy formed out of gas from within the protogalaxy. However, the relative specific frequencies in the MPC and MRC components would be the same in any other scheme; they depend only on the assumption that the MPC clusters and metal-poor halo light go together, and that the MRC clusters and metal-rich halo light go together.

Several other galaxies appear to present a story with strong similarities to that in NGC 4472, such as NGC 1399 (Ostrov et al. 1998; Forbes et al. 1998) and M87 itself (Whitmore et al. 1995; Kundu et al. 1999). Two major subgroups dominate the MDF, each of which displays little or no metallicity gradient in itself. The MRC is more centrally concentrated, giving rise to a net [Fe/H] gradient in the whole GCS. Conversely, in galaxies with clearly unimodal MDFs, none so far show any clear evidence for metallicity gradients. In summary, the presence or absence of gradients in halo metallicity appears to connect strongly with the form of the MDF. Each separate stage of cluster formation generated clusters at similar metallicities all across the potential well of the galaxy, and it is only the different radial concentrations of these components that gives rise to an overall gradient in the total GCS.

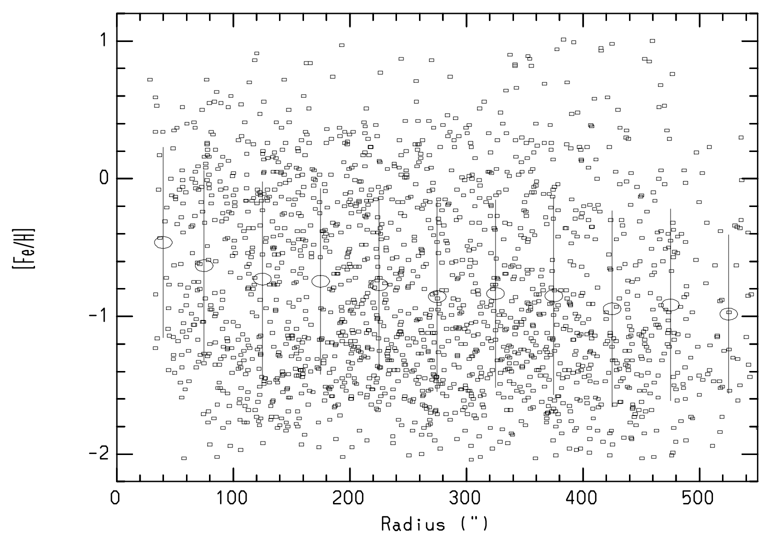

Fig. 1.46. Metallicity vs. radius for globular clusters in NGC 4472, from Geisler et al. (1996). Note the bimodal distribution in metallicity, with the redder (more metal-rich) population more centrally concentrated. Figure courtesy Dr. D. Geisler

Another elliptical galaxy of special interest is NGC 5128, the dominant galaxy in the small, nearby Centaurus group ($d = 3.9$ Mpc). The importance of this galaxy is that it is the only giant elliptical in which we have been able to directly compare the MDF of the *halo stars* with the *clusters*. G. Harris et al. (1999) have used deep HST/WFPC2 photometry in V and I to obtain direct color-magnitude photometry of the red-giant stars in the outer halo of NGC 5128, from which they generate an MDF by interpolation within standard RGB evolutionary tracks. The comparison between the two MDFs (clusters and halo stars) is shown in Fig. 1.48.

The NGC 5128 halo stars display an MDF with at least two major components; roughly two-thirds of the stars are in the narrow metal-rich component located at [Fe/H](peak)$= -0.3$ and with dispersion σ[Fe/H]$\simeq 0.25$. Remarkably, the metal-rich part of the bimodal cluster MDF has the same location and the same dispersion. Its specific frequency (that is, the ratio of MRC clusters to MRC stars) is $S_N(MRC) \simeq 1.5$. By contrast, the metal-poor component makes up about a third of the halo stars but about two-thirds of the clusters, so that its specific frequency is $S_N(MPC) \simeq 4.3$. This is, however, only a local estimate for one spot in the halo. The global value of S_N(MPC) across the entire galaxy would be larger if the inner parts of the halo contain relatively more MRC stars; that is, if the halo has a mean metallicity gradient.

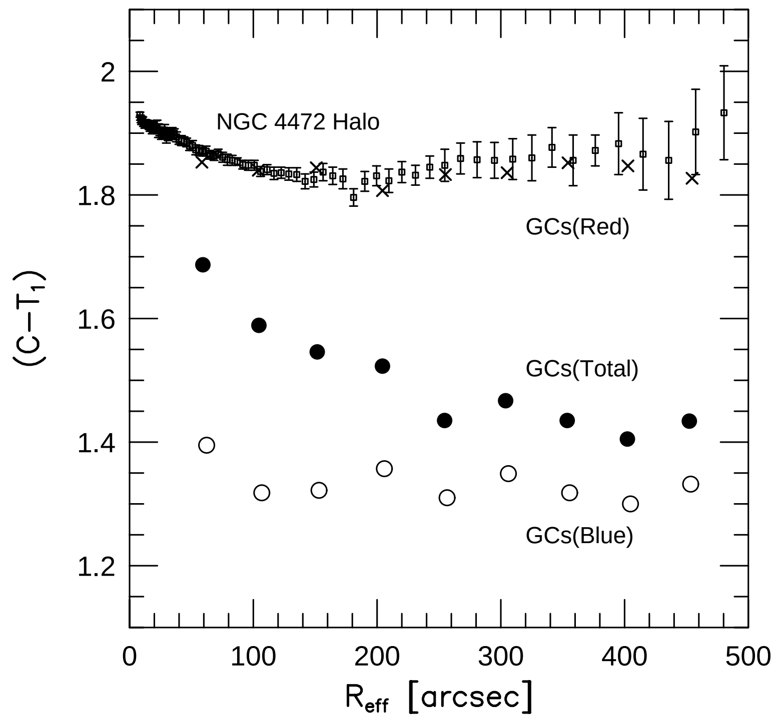

Fig. 1.47. Mean color $(C - T_1)$ vs. radius for globular clusters in NGC 4472, from Lee et al. (1998). Plotted separately are the red (metal-rich) globular clusters as crosses; blue (metal-poor) clusters as open circles; and the integrated color of the NGC 4472 halo light (small dots with error bars). The mean color of all clusters combined (red + blue) is shown as the solid dots. Figure courtesy Dr. D. Geisler

G. Harris et al. (1999) argue that the most likely interpretation of the early history of this galaxy is an *in situ* formation model much like the one outlined above: two rather distinct stages of star formation, in which the first (metal-poor) one left most of the gas unconverted, but slightly enriched from the first, low-efficiency round of star formation. The later (metal-richer) burst then converted most of the gas and produced the main visible bulk of the galaxy. Though later accretions of small satellites must have played some role in building up NGC 5128 (one gas-rich accretion has clearly occurred recently to fuel the starburst activity within the inner ~ 5 kpc), these do not seem to have affected the outer-halo regions.

NGC 5128 may of course not be typical of all ellipticals. But interactions of the type it is now undergoing are now realized to be fairly commonplace for large galaxies, so there is every reason for optimism that we can use it to learn about the early evolution of many giant ellipticals.

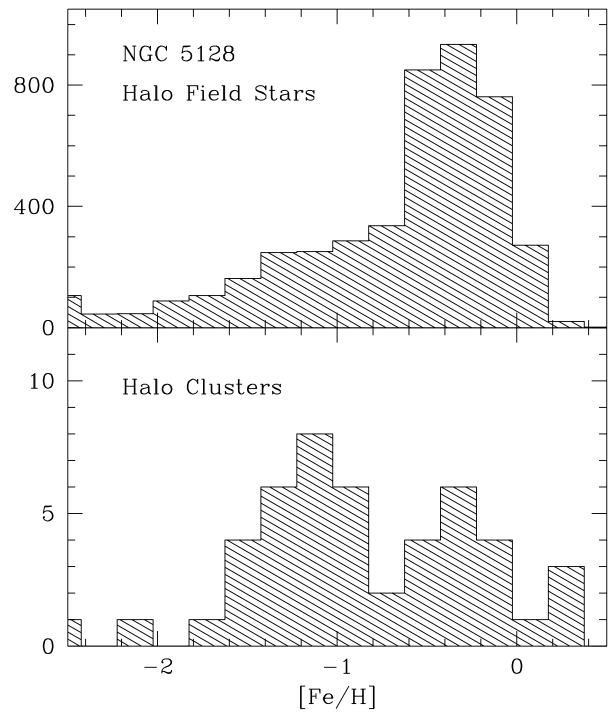

Fig. 1.48. *Upper panel:* Metallicity distribution function for red-giant stars in the outer halo of NGC 5128, at a projected location 20 kpc from the galaxy center. *Lower panel:* MDF for the globular clusters in the halo of NGC 5128 more distant than $4'$ (4.5 kpc) from the galaxy center. Data are from G. Harris et al. (1992, 1999)

The analysis of GCS metallicity distributions has been one of the most productive routes to understanding cluster formation and the early histories of galaxies. I urge interested readers to see the extensive discussion of Ashman & Zepf (1998) for more of the history of MDFs and their analysis.

1.6.5 Radial Velocities and Dynamics

If we want to study the dynamics of the halo in a distant galaxy, then globular clusters give us the same advantage over halo field stars as they did for the metallicity distributions: because we can identify them one by one, we can build up the actual velocity *distribution function* rather than just a luminosity-weighted mean. Potentially, we can use cluster velocities in remote galaxies to determine (a) the kinematic differences between MPC and MRC clusters, where they are present; (b) the mass distribution $M(r)$ and the amount of dark matter; (c) the orbital distribution and the degree of anisotropy; and (d) the presence (or absence) of "intergalactic" globular clusters, i.e. clusters moving freely in the potential well of the galaxy cluster as a whole.

Obtaining the necessary velocity measurements is a demanding job, requiring the biggest available optical telescopes and *large* samples of clusters. Early velocity measurements were accomplished for a few dozens of clusters in three giant ellipticals, M87 (Mould et al. 1987, 1990; Huchra & Brodie 1987), NGC 4472 (Mould et al. 1990), and NGC 5128 (H.Harris et al. 1988). These studies were consistent with the expected results that the velocity distributions were roughly isotropic and that the velocity dispersion was nearly uniform with radius, thus $M(r) \sim r$. However, more recent studies – with higher quality data and significantly larger samples – have begun to reveal more interesting features. For M87, Cohen & Ryzhov (1997) have used a sample of ~ 200 clusters extending out to $r \sim 30$ kpc to suggest that the velocity dispersion *rises* with radius, indicating $M(r) \sim r^{1.8}$. The cluster velocities thus suggest the presence of an extensive amount of halo mass which bridges the mass profile of the central cD galaxy to the larger-scale mass distribution as determined from the hot X-ray gas on 100-kpc scales. Still larger samples of cluster velocities for M87 are in progress, and may be able to provide first hints on the velocity anisotropy parameters.

Sharples et al. (1998) have published the first stages of a study of similar scale on NGC 4472, the other Virgo supergiant. Other notable studies for disk galaxies include the recent work on the Sombrero Sa galaxy (NGC 4594) by Bridges et al. (1997) and on NGC 3115 by Kavelaars (1998). For NGC 1399, the central cD galaxy in Fornax, several dozen cluster velocities have now been obtained (Grillmair et al. 1994; Minniti et al. 1998; Kissler-Patig 1998; Kissler-Patig et al. 1999). They find that the GCS velocity dispersion at $r \gtrsim 20$ kpc is noticeably higher than that of the inner halo stars or clusters (as deduced from the integrated light and planetary nebulae) but similar instead to the population of *galaxies* around NGC 1399, suggesting that many of the globular clusters in the cD envelope may belong to the Fornax potential as a whole rather than the central elliptical. Some contamination from neighboring ellipticals is also a possibility, and considerably more datapoints will be needed to sort out the alternatives (see Kissler-Patig et al. 1999).

Here we end our overview of globular cluster systems in different galaxies. After a brief detour into the Hubble constant (next section), we will return in the last two sections to a discussion of current ideas about globular cluster formation and the early history of galaxies.

1.7 THE GCLF AND THE HUBBLE CONSTANT

The only goal of science is the diminution of the distance between present knowledge and truth.

Steven Goldberg

Globular cluster systems are astrophysically most important for what they can tell us about galaxy formation. Confronted with the rich variety of observational information we now have for GCSs in many galaxies, and the range of implications it all has for galaxy formation, it is somewhat surprising to recall that they were historically first regarded as attractive for their potential as *extragalactic distance indicators* – that is, standard candles. In this section, we will take a brief look at the history of attempts to use globular clusters as standard candles; discuss the basic technique in its contemporary form; work through the empirical calibration issues; and finally, see how it is applied to remote galaxies and derive a new estimate of H_0.

1.7.1 Origins

The brightest globular clusters are luminous ($M_V \lesssim -11$) and thus detectable at distances far beyond the Local Group – particularly in giant ellipticals with populous GCSs that fill up the bright end of the cluster luminosity distribution. M87, the central cD in the Virgo cluster, was the first such galaxy to attract attention. Attempts to use the brightest clusters began with the discovery paper by Baum (1955), who first noted the presence of globular clusters around M87 visible on deep photographic plates. In several later papers (Sandage 1968; Racine 1968; van den Bergh 1969; de Vaucouleurs 1970; Hodge 1974), the mean magnitudes of these few brightest clusters were used to estimate the distance to M87, under the assumption that their intrinsic luminosities were the same as those of Mayall II (the brightest cluster in M31), or ω Centauri (the brightest in the Milky Way), or some average of the most luminous clusters in the Local Group galaxies. All of these attempts were eventually abandoned after it became clear that the brightest clusters drawn from a huge statistical sample – like the M87 GCS – would be more luminous than those drawn from the much smaller Milky Way and M31 samples, even if their GCLFs were basically similar (which was itself an unproven assumption).

The modern approach to employing the GCLF begins with the work of Hanes (1977), who carried out a large photographic survey of the globular cluster systems in several Virgo ellipticals. The photometric limits of this material still fell well short of the GCLF turnover, but the basic principle was established that *the entire GCLF* had considerably more information than just its bright tip, and could be matched in its entirety with the calibrating GCSs in the Milky Way or M31. The Gaussian interpolation model for the GCLF was also employed in essentially the same way we use it today. With the benefit of hindsight (see the discussion of Harris 1988b), we can see from Hanes' analysis that he would have correctly predicted the Virgo GCLF turnover magnitude if he had known the right value of the GCLF dispersion σ for these ellipticals. Somewhat deeper photographic photometry for additional Virgo ellipticals was obtained by Strom et al. (1981) and Forte et al. (1981), with similar results.

The subject – like most other areas of observational astronomy – was revolutionized by the deployment of the enormously more sensitive CCD cameras, beginning in the mid-1980's. At last, the anticipated GCLF turnover was believed to be within reach of the new CCD cameras on large telescopes. Once again, M87 was the first target: long exposures with a first-generation CCD camera by van den Bergh et al. (1985) attained a photometric limit of $B \simeq 25.4$. They did indeed reach the turnover point, though they could not definitively prove it, since the photometric limit lay *just* past the putative turnover. Still deeper $B-$band data were obtained by Harris et al. (1991) for three other Virgo ellipticals, which finally revealed that the turnover had been reached and passed, with the data exhibiting a clearly visible downturn extending 1.5 mag past the peak. For the first time, it was possible to argue on strictly observational grounds that the GCLF had the same fundamental shape in E galaxies as in the Milky Way and M31. With the advent of the Hubble Space Telescope era in the 1990's, considerably more distant targets have come within reach, extending to distances where galactic motions are presumed to be dominated by the cosmological Hubble flow and peculiar motions are negligible.

1.7.2 The Method: Operating Principles

In its modern form, the GCLF is the simplest of standard candles that apply to remote galaxies. For the purposes of this section, we will use the classic Gaussian-like form of the luminosity distribution (number of clusters per unit magnitude). The observational goal is nothing more than *to find the apparent magnitude V^0 of the turnover point.* Once an absolute magnitude M_V^0 is assumed, the distance modulus follows immediately. The precepts of the technique are laid out in Secker & Harris (1993) and in the reviews of Jacoby et al. (1992) and Whitmore (1997). Briefly, the basic attractions of the GCLF method are as follows:

- M_V^0 is more luminous than any other stellar standard candle except for supernovae. With the HST cameras ($V(lim) \gtrsim 28$), its range extends to $d \sim 120$ Mpc and potentially further.
- Globular clusters are old-halo objects, so in other galaxies they are as free as possible from problems associated with dust and reddening inside the target galaxy.
- They are nonvariable objects, thus straightforward to measure (no repeat observations are necessary).
- They are most numerous in giant E galaxies which reside at the centers of rich galaxy clusters. These same objects are the ones which are the main landmarks in the Hubble flow, thus concerns about peculiar motions or interloping galaxies are minimized.

Clearly, it shares at least some of these advantages with other techniques based on old stellar populations that work at somewhat shorter range: the planetary nebula luminosity function, surface brightness fluctuations, and the RGB tip luminosity (Jacoby et al. 1992; Lee et al. 1993a).

Having listed its attractions, we must also be careful to state the concerns and potential pitfalls. There are two obvious worries arising from the astrophysical side. *First*, globular clusters are small *stellar systems* rather than individual stars. We cannot predict their luminosities starting from a secure basis in stellar physics, as we can do for (e.g.) Cepheids, planetary nebulae, or RGB tip stars. Indeed, to predict the luminosity distribution of globular clusters, we would first have to know a great deal about how they form. But understanding their formation process almost certainly involves complex, messy gas dynamics (see Section 8 below), and at the moment, we have no such complete theory on hand. In any case, we might well expect *a priori* that clusters would form with different typical masses or mass distributions in different environments, such as at different locations within one galaxy, or between galaxies of widely different types.

Second, globular clusters are $\sim 10-15$-Gyr-old objects, and as such they have been subjected to a Hubble time's worth of dynamical erosion within the tidal fields of their parent galaxies. Since the efficiencies of these erosive processes also depend on environment (Section 5), shouldn't we expect the GCLFs to have evolved into different shapes or mean luminosities in different galaxies, even if they started out the same?

In the absence of direct observations, these theoretical expectations seem formidable. But we should not mistake the relative roles of theory and experiment: that is, arguments based on whatever is the current state of theory should not prevent us from going out and discovering what the real objects are like. For the distance scale, the fundamental issue (Jacoby et al. 1992) can be simply stated: *Any standard candle must be calibrated strictly on observational grounds; the role of theory is to explain what we actually see.* Theory may give us an initial motivation or overall physical understanding of a particular standard candle, but the only way that our carefully constructed distance scale can be independent of changes in the astrophysical models is to build it purely on measurement.

At the same time, we must recognize the challenges as honestly as we can. If we are to use the GCLF as a standard candle, we must have clear evidence that the turnover magnitude M_V^0 is in fact the same from galaxy to galaxy.

More precisely, we must be confident that the behavior of M_V^0 is *repeatable* from galaxy to galaxy. This is, of course, not a black-and-white statement but rather a matter of degree: like any other empirical standard candle, M_V^0 cannot be a perfect, ideally uniform number. But is it a "constant" at the level of, say, ± 0.1 magnitude? ± 0.2 mag? or worse? This is the practical question which determines how interesting the GCLF actually is as a distance indicator, and which must be settled empirically.

1.7.3 Calibration

We calibrate the turnover luminosity M_V^0 by measuring it in several other nearby galaxies whose distances are well established from precise stellar standard candles. But just using the Milky Way and M31 (Section 5) will not do. We will be particularly interested in using the GCLF in remote giant ellipticals, and these are galaxies of quite a different type than our nearby spirals.

The closest large collections of E galaxies are in the Virgo and Fornax clusters. Fortunately, these are near enough that their distances can be measured through a variety of stellar standard candles, and so these two clusters must be our main proving grounds for the GCLF calibration. Here, I will use galaxy distances established from four different methods which have sound physical bases and plausible claims to precisions approaching ± 0.1 magnitude in distance modulus: (a) the period-luminosity relation for Cepheids; (b) the luminosity function for planetary nebulae (PNLF); (c) surface brightness fluctuations for old-halo stellar populations (SBF); and (d) the red-giant branch tip luminosity (TRGB). For extensive discussions of these (and other) methods, see Jacoby et al. (1992) and Lee et al. (1993a). In Table 1.10, recent results from these four methods are listed for several galaxy groups and individual galaxies with globular cluster systems. In most cases, the mutual agreements among these methods are good, and bear out their claimed accuracies in the references listed.

The final column of the table gives the adopted mean distance modulus for each group, along with the *internal* r.m.s. uncertainty of the mean. As a gauge of the true (external) uncertainty, we can note that to within ± 0.1 in distance modulus, the absolute zeropoints of each technique are consistent with the Local Group (LMC and M31) distance scale discussed in Section 2 above.

Next, we need to have well established *apparent magnitudes* V^0 for the GCLF turnover levels in as many galaxies as possible. The most straightforward numerical technique is to start with the observed GCLF (corrected for background contamination and photometric incompleteness; see the Appendix) and fit any of the adopted interpolation functions to it – usually the Gaussian, but others such as the t_5 function have been used too. The best-fit function gives the nominal apparent magnitude V^0 of the turnover point. Secker & Harris (1993) define a more advanced maximum-likelihood procedure for fitting the raw data (that is, the list of detected objects in the field, sorted by magnitude) to the adopted function, convolved with the photometric error and completeness functions and added to the observed background LF. Both approaches have proved to generate valid results, though the latter method provides a more rigorous understanding of the internal uncertainties.

To determine V^0, we need to have GCLF photometry extending clearly past the turnover: the deeper the limit, the more precisely we can identify it independent of assumptions about the shape or dispersion of the GCLF as a whole. (It is important to note here that we do *not* necessarily want to

Table 1.10. Distance moduli for nearby galaxy groups

Galaxy Group	$(m-M)_0$	Method	Sources	Mean
Virgo Cluster	30.99 ± 0.08	Cepheids	1,2,3,4	30.97 ± 0.04
	30.98 ± 0.18	TRGB	5	
	30.84 ± 0.08	PNLF	6,7	
	31.02 ± 0.05	SBF	8,9,10,11	
Fornax Cluster	31.35 ± 0.07	Cepheids	12	31.27 ± 0.04
	31.14 ± 0.14	PNLF	13	
	31.23 ± 0.06	SBF	8	
Leo I Group	30.01 ± 0.19	Cepheids	14	30.17 ± 0.05
	30.30 ± 0.28	TRGB	15	
	30.10 ± 0.08	PNLF	16,17	
	30.20 ± 0.05	SBF	8,11,18	
Coma I Group	30.08 ± 0.08	PNLF	19	30.08 ± 0.07
	30.08 ± 0.07	SBF	20	
Coma II Group	30.54 ± 0.05	PNLF	19	30.81 ± 0.14
	30.95 ± 0.07	SBF	8,20	
NGC 4365	31.73 ± 0.10	SBF	8	31.73 ± 0.10
NGC 3115	30.29 ± 0.20	TRGB	21	30.16 ± 0.10
	30.17 ± 0.13	PNLF	22	
	29.9 ± 0.25	SBF	21,22	
NGC 4594	29.74 ± 0.14	PNLF	23	29.70 ± 0.10
	29.66 ± 0.08	SBF	22	

Sources: (1) Ferrarese et al. 1996 (2) Pierce et al. 1994 (3) Saha et al. 1996a (4) Saha et al. 1996b (5) Harris et al. 1998b (6) Jacoby et al. 1990 (7) Ciardullo et al. 1998 (8) Tonry et al. 1997 (9) Neilsen et al. 1997 (10) Pahre & Mould 1994 (11) Morris & Shanks 1998 (12) Madore et al. 1998 (13) McMillan et al. 1993 (14) Graham et al. 1997 (15) Sakai et al. 1997 (16) Ciardullo et al. 1989 (17) Feldmeier et al. 1997 (18) Sodemann & Thomsen 1996 (19) Jacoby et al. 1996 (20) Simard & Pritchet 1994 (21) Kundu & Whitmore 1998 (22) Ciardullo et al. 1993 (23) Ford et al. 1996 use a fitting function which will match the entire GCLF, which may or may not be asymmetric at magnitudes far out in the wings. The entire goal of the numerical exercise is *to estimate the magnitude of the turnover point as accurately as possible*; thus, we want a fitting function which will describe the peak area of the GCLF accurately and simply. In other words, it is to our advantage to use a simple, robust function which will not be overly sensitive to the behavior of the GCLF in the far wings. The Gaussian and t_5 functions, with just two free parameters, meet these requirements well.)

The results for E galaxies with well determined GCLF turnovers are listed in Tables 1.11 and 1.12, while Table 1.13 gives the same results for several disk galaxies. (Note that the turnover luminosities for the Milky Way and the Local Group dE's are already converted to absolute magnitude.) The fourth column in each table gives the magnitude limit of the photometry relative to the turnover level; obviously, the larger this quantity is, the more well determined the turnover point will be. The absolute magnitude of the turnover in each galaxy is obtained by subtraction of the intrinsic distance moduli in Table 1.10, and subtraction of the foreground absorption A_V. Fortunately, A_V is small in most cases, since almost all the galaxies listed here are at high latitude.

The values of M_V^0 in the individual galaxies are shown in Figs. 1.49 and 1.50, and the mean values are listed in Table 1.14.

The results for the giant ellipticals are particularly important, since these act as our calibrators for the more remote targets. From the first entry in Table 1.14, we see that the gE galaxy-to-galaxy scatter in M_V^0 is at the level of ± 0.15 mag *without any further corrections* due to environment, metallicity, luminosity, or other possible parameters. Much the same scatter emerges if we use only the gE galaxies within one cluster (Virgo or Fornax) where they are all at a common distance (cf. Harris et al. 1991; Jacoby et al. 1992; Whitmore 1997 for similar discussions). This all-important quantity determines the intrinsic accuracy that we can expect from the technique. Clearly, part of the dispersion in M_V^0 must be due simply to the statistical uncertainty in determining the apparent magnitude of the turnover from the observed GCLF (which is typically ± 0.1 mag at best; see below), and part must be due to uncertainties in the adopted distances to the calibrating galaxies (which again are likely to be ± 0.1 mag at best). When these factors are taken into account, the raw observed scatter in the turnover magnitudes is encouragingly small.

In summary, I suggest that the directly observed dispersion in the turnover luminosity gives a reasonable estimate of the precision we can expect from the technique: for giant E galaxies with well populated GCLFs, the expected uncertainty in the resulting distance modulus is near ± 0.15 mag.

One remaining anomaly within the set of gE galaxies is a slight systematic discrepancy between the Fornax and Virgo subsamples. For the six Virgo ellipticals by themselves, we have $\langle M_V^0 \rangle \simeq -7.26 \pm 0.07$, while for the six Fornax ellipticals, $\langle M_V^0 \rangle \simeq -7.47 \pm 0.07$. These differ formally by (0.21 ± 0.10),

Table 1.11. GCLF turnover magnitudes for giant E galaxies

Galaxy Group	Galaxy	V^0(turnover)	$V(lim) - V^0$	Sources
Virgo Cluster	N4472	23.87 ± 0.07	$\simeq 1.5$	1,2,3
	N4478	23.82 ± 0.38	2.7	4
	N4486	23.71 ± 0.04	$\simeq 2.1$	5,6,7,8,9
	N4552	23.70 ± 0.30	0.7	2
	N4649	23.66 ± 0.10	$\simeq 1.8$	1
	N4697	23.50 ± 0.20	1.2	10
Fornax Cluster	N1344	23.80 ± 0.25	1.0	11
	N1374	23.52 ± 0.14	0.5	12
	N1379	23.92 ± 0.20	1.0	12,13
	N1399	23.86 ± 0.06	1.0	11,12,14,15
	N1404	23.94 ± 0.08	1.0	11,15,16
	N1427	23.78 ± 0.21	-0.2	12
Leo I Group	N3377	22.95 ± 0.54	1.3	17
	N3379	22.41 ± 0.42	1.3	17
Coma I Group	N4278	23.23 ± 0.11	1.6	18
Coma II Group	N4494	23.34 ± 0.18	1.7	18,19
NGC 4365	N4365	24.42 ± 0.18	0.8	1,2,20

Sources: (1) Secker & Harris 1993 (2) Ajhar et al. 1994 (3) Lee et al. 1998 (4) Neilsen et al. 1997 (5) Harris et al. 1991 (6) McLaughlin et al. 1994 (7) Whitmore et al. 1995 (8) Harris et al. 1998a (9) Kundu et al. 1999 (10) Kavelaars & Gladman 1998 (11) Blakeslee & Tonry 1996 (12) Kohle et al. 1996 (13) Elson et al. 1998 (14) Bridges et al. 1991 (15) Grillmair et al. 1999 (16) Richtler et al. 1992 (17) Harris 1990b (18) Forbes 1996b (19) Fleming et al. 1995 (20) Forbes 1996a

Table 1.12. GCLF turnover magnitudes for dwarf E galaxies

Galaxy Group	Galaxy	V^0(turnover)	$V(lim) - V^0$	Sources
Virgo Cluster	8 dE's	24.1 ± 0.3	0.7	1
NGC 3115	DW1	23.1 ± 0.3	1.4	2
Local Group	NGC 147	-5.99 ± 0.92	2:	3
	NGC 185	-6.49 ± 0.71	2:	3
	NGC 205	-7.27 ± 0.27	2:	3
	Fornax	-7.06 ± 0.95	3:	3
	Sagittarius	-6.28 ± 1.21	4:	4

Sources: (1) Durrell et al. 1996a (2) Durrell et al. 1996b (3) Harris 1991 (4) This paper

Table 1.13. GCLF turnover magnitudes for disk galaxies

Group	Galaxy	Type	V^0(turnover)	$V(lim) - V^0$	Sources
Fornax	N1380	S0	23.92 ± 0.20	1.1	1,2
NGC 3115	N3115	S0	22.37 ± 0.05	0.7	3
Virgo SE	N4594	Sa	23.3 ± 0.3	1.0	4
Coma I:	N4565	Sb	22.63 ± 0.21	0.8	5
M81	M81	Sb	20.30 ± 0.3	2	6
Local Group	M31	Sb	17.00 ± 0.12	2	7,8
Local Group	Milky Way	Sbc	-7.68 ± 0.14	5	9
Local Group	M33	Sc	17.74 ± 0.17	2	10
Local Group	LMC	Im	11.13 ± 0.32	3	10

Sources: (1) Blakeslee & Tonry 1996 (2) Kissler-Patig et al. 1997b (3) Kundu & Whitmore 1998 (4) Bridges & Hanes 1992 (5) Fleming et al. 1995 (6) Perelmuter & Racine 1995 (7) Reed et al. 1994 (8) Secker 1992 (9) This paper (Chp 5) (10) Harris 1991 significant at the two-standard-deviation level. Why? If the GCLF turnover is, indeed, fundamentally similar in these rich-cluster ellipticals and subject only to *random* differences from one galaxy to another, then this discrepancy would suggest that we have either overestimated the distance to Fornax, or underestimated the distance to Virgo, or some combination of both. But the stellar standard candles listed above agree quite well with one another in each cluster. This puzzling discrepancy is not large; but it suggests, perhaps, that the external uncertainty in the GCLF turnover method may be closer to ± 0.2 mag.

What of the other types of galaxies? From the evidence so far, dwarf ellipticals have turnover luminosities that are *fainter* by $\sim 0.3 - 0.4$ mag in M_V than in the giants. Quite obviously, though, measuring the turnover in any one dwarf is a risky business because of the small sample size (perhaps only one or two dozen globular clusters per galaxy even in the best cases; see Durrell et al. 1996a,b). Many must be averaged together to beat down the individual statistical uncertainties.

In the disk galaxies, the turnover may be slightly *brighter* (by $\sim$ 0.2 mag) than in gE's, though the nominal difference is not strongly significant. This latter result, if real, may be tangible evidence that dynamical evolution of globular clusters in disk galaxies has been somewhat stronger due to disk shocking, which would remove a higher proportion of the fainter clusters. The one strikingly anomalous case is NGC 4594, with a much fainter turnover level than average. Although its distance seems relatively well determined (PNLF, SBF), the GCLF turnover magnitude relies on only one small-field CCD study and may be suspect. This galaxy is the nearest giant edge-on Sa and should be studied in much more detail.

Table 1.14. Final GCLF turnover luminosities

Galaxy Type	N	Mean M_V^0	rms scatter
Giant Ellipticals	16	-7.33 ± 0.04	0.15
Dwarf Ellipticals	14	-6.90 ± 0.17	0.6:
All Disk Galaxies	9	-7.46 ± 0.08	0.22
S0 and Sb	6	-7.57 ± 0.08	0.20

NB: The mean for the giant ellipticals excludes NGC 4278, at $M_V^0 = -6.9$; its distance modulus is probably suspect.

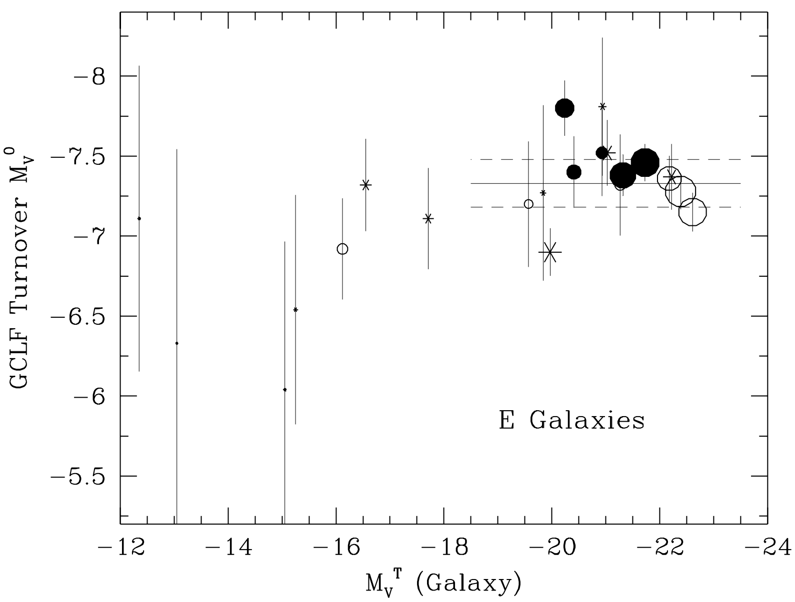

Fig. 1.49. GCLF turnover luminosity M_V^0 for elliptical galaxies, plotted against galaxy luminosity M_V^T. *Solid dots* are ellipticals in the Fornax cluster, *open circles* are Virgo ellipticals, and *asterisks* are ellipticals in smaller groups. Symbol size goes in inverse proportion to the internal uncertainty in the turnover (smaller symbols have larger random errors). The horizontal solid line indicates the mean $\langle M_V^0 \rangle$ for the giant ellipticals, with the ± 0.15 galaxy-to-galaxy range indicated by the dashed lines

1.7.4 Functional Fitting and the Role of the Dispersion

We see that the absolute magnitude of the turnover point is reasonably similar in widely different galaxies. Now, what can we say about the *dispersion* of the GCLF? Specifically, in our Gaussian interpolation model, is the standard deviation σ_G reasonably similar from one galaxy to another?

An important side note here is that in practice, σ_G really represents the shape of the *bright half* of the GCLF, since in most galaxies beyond the Local Group we do not have data that extend much beyond the turnover point itself. Thus if the relative numbers of faint clusters were to differ wildly from one type of galaxy to another, we would not yet have any way to see it (nor would it matter for the standard-candle calibration). However, to test the uniformity of σ_G, we want to use only the calibrating galaxies for which the limit of the photometry is clearly fainter than the turnover. If the data fall short of V^0, or just barely reach it, then it is generally not possible to fit a Gaussian curve to the data and solve *simultaneously* for both V^0 and σ_G; the two parameters are correlated, and their error bars are asymmetric.

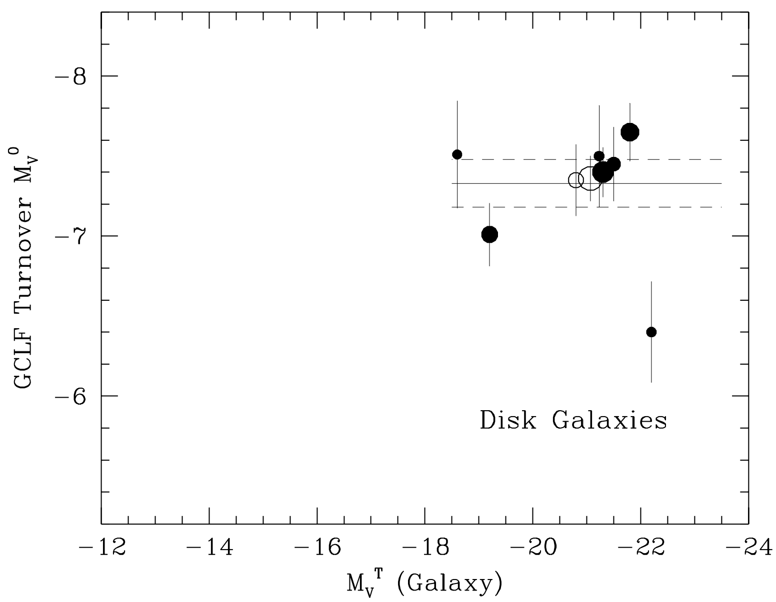

Fig. 1.50. GCLF turnover luminosity M_V^0 for disk galaxies. *Solid dots* are spirals (Sa to Im types) and *open circles* are S0's. The horizontal lines are taken from the previous figure, and indicate the mean and standard deviation for giant ellipticals. Most of the large disk galaxie sit slightly above the mean line for the ellipticals. The anomalously low point is the Sa galaxy NGC 4594

This latter numerical problem was already realized in attempts to fit the first deep CCD data in M87 (van den Bergh et al. 1985; Hanes & Whittaker 1987), and is also discussed at length in Harris (1988b) and Secker & Harris (1993). The reason for the asymmetry can be seen immediately if we refer again to Fig. 1.43: if the observations do not extend past the turnover, then there are no faint-end data points to constrain the upper limits on either σ_G or V^0, and a statistically good fit can be obtained by choices of these parameters that may be much larger than the true values. By contrast, values that are much too *small* are ruled out by the well determined bright-end observations. The net result is unfortunately that *both the dispersion and the turnover tend to be overestimated* if both are allowed to float in the fitted solution.

Best-fit values of σ_G are listed in Table 1.15 for most of the same galaxies listed above. Ellipticals are listed on the left, and disk galaxies on the right. For the six disk galaxies, the weighted mean is $\langle\sigma_G\rangle = 1.21 \pm 0.05$. For 12 ellipticals (excluding NGC 4478, which is a peculiar tidally truncated companion of NGC 4472), we obtain $\langle\sigma_G\rangle = 1.36 \pm 0.03$.

An interesting comparison of this mean value can be obtained from the results of GCLFs in 14 BCG galaxies from the surface brightness fluctuation

Table 1.15. GCLF dispersion measurements

Ellipticals		Disks	
Galaxy	σ_G	Galaxy	σ_G
N1344	1.35 ± 0.18	Milky Way	1.15 ± 0.10
N1379	1.55 ± 0.21	M31	1.06 ± 0.10
N1399	1.38 ± 0.09	M33	1.2 :
N1404	1.32 ± 0.14	N1380	1.30 ± 0.17
N4278	1.21 ± 0.09	N3115	1.29 ± 0.06
N4365	1.49 ± 0.20	N4565	1.35 ± 0.22
N4472	1.47 ± 0.08		
N4478	1.16 ± 0.21		
N4486	1.40 ± 0.06		
N4494	1.09 ± 0.11		
N4636	1.35 ± 0.06		
N4649	1.26 ± 0.08		
N5846*	1.34 ± 0.06		

*Source for NGC 5846: Forbes et al. 1996a

study of Blakeslee et al. (1997). They find $\langle\sigma_G\rangle = 1.43 \pm 0.06$. (In their SBF analysis, only the few brightest globular clusters are actually resolved on the raw images, but the fluctuation contribution due to the fainter unresolved ones must be numerically removed before the fluctuation signal from the halo light can be determined. They assume that the GCLF follows a Gaussian shape with an assumed M_V^0 equal to that of M87, and then solve for the dispersion.)

In summary, a mean value $\sigma_G = 1.4 \pm 0.05$ appears to match most giant ellipticals rather well, and $\sigma_G = 1.2 \pm 0.05$ will match most spirals.

Some common-sense prescriptions can now be written down for the actual business of fitting an interpolation function to an observed GCLF. Starting with the observations of cluster numbers vs. magnitude, your goal is simply to *estimate the turnover point as accurately as possible.* Choose a simple, robust interpolation function which will match the center of the distribution and don't worry about the extreme wings. But should you try to solve for both V^0 and σ_G, which are the two free parameters in the function? This depends completely on how deep your photometry reaches. Experience shows that if you have fully corrected your raw data for photometric incompleteness and subtracted off the contaminating background LF, and you clearly see that your data reach *a magnitude or more past the turnover point*, then you can safely fit one of the recommended functions (Gaussian or t_5) to it and solve for both parameters. However, if your photometric limit falls short of the turnover, or does not go *clearly* past it, then your best course of action is to *assume* a value for the dispersion and solve only for the turnover magnitude. This approach will introduce some additional random uncertainty in V^0, but will considerably reduce its systematic uncertainty.

The actual function fitting process can be developed into one in which the assumed model (Gaussian or t_5) is convolved with the photometric completeness and measurement uncertainty functions (see the Appendix), added to the background LF, and then matched to the raw, uncorrected LF. A maximum-likelihood implementation of this approach is described in Secker & Harris (1993).

Putting these results together, we now have some confidence *on strictly empirical grounds* that the turnover luminosity in gE galaxies has an observational scatter near ± 0.15 mag, and a Gaussian dispersion $\sigma_G \simeq 1.4 \pm 0.05$. These statements apply to the central cD-type galaxies in Virgo and Fornax, as well as to other gE's in many groups and clusters. The GCLFs in dwarf ellipticals and in disk galaxies are noticeably, but not radically, different in mean luminosity and dispersion.

This is all the evidence we need to begin using the GCLF as a standard candle for more remote ellipticals. The near-uniformity of the GCLF luminosity and shape, in an enormous range of galaxies, is a surprising phenomenon on astrophysical grounds, and is one of the most remarkable and fundamentally important characteristics of globular cluster systems.

1.7.5 The Hubble Constant

To measure H_0, we need GCLF measurements in some target galaxies that are much more distant than our main group of calibrators in Virgo and Fornax. Such observations are still a bit scarce, but the numbers are steadily growing. Our preferred route will be the classic one through the "Hubble diagram". We start with Hubble's law for redshift v_r and distance d:

$$v_r = H_0 \, d \tag{1.52}$$

or in magnitude form where d is measured in Mpc and v_r in km s^{-1},

$$5 \log v_r \;=\; 5 \log H_0 \;+\; (m-M)_0 \;-\; 25 \tag{1.53}$$

Now substitute the apparent magnitude of the GCLF turnover, $V^0 = M_V^0 + (m-M)_0$, and we obtain

$$\log v_r \;=\; 0.2V^0 \;+\; \log H_0 \;-\; 0.2M_V^0 \;-\; 5 \tag{1.54}$$

Thus a plot of (log v_r) against apparent magnitude V^0 for a sample of giant elliptical galaxies should define a straight line of slope 0.2. The zeropoint (intercept) is given by $\langle \log v_r - 0.2V^0 \rangle = \log H_0 - 0.2M_V^0 - 5$, where the mean in brackets is taken over the set of observed data points. Once we insert our adopted value of M_V^0, the value of the Hubble constant H_0 follows immediately.

Relevant data for a total of 10 galaxies or groups ranging from the Virgo cluster out to the Coma cluster (the most remote system in which the GCLF turnover has been detected) are listed in Table 1.16 and plotted in Fig. 1.51. This figure is the first published "Hubble diagram" based on globular cluster luminosities, and it has been made possible above all by the recent HST photometry of a few remote ellipticals.

In the Table, the entries for Virgo and Fornax are the mean $\langle V^0 \rangle$ values taken from Table 1.11 above. The cosmological recession velocities $v_r = cz$ for each target assume a Local Group infall to Virgo of 250 ± 100 km s^{-1} (e.g., Ford et al. 1996; Hamuy et al. 1996; Jerjen & Tammann 1993, among many others). For the mean radial velocities of the clusters, especially Virgo and Fornax, see the discussions of Colless & Dunn (1996), Girardi et al. (1993), Huchra (1988), Binggeli et al. (1993), Mould et al. (1995), and Hamuy et al. (1996). The Coma cluster ellipticals (IC 4051 and NGC 4874, and the lower limit for NGC 4881) provide especially strong leverage on the result for H_0, since they are easily the most distant ones in the list, and the correction of the cluster velocity to the cosmological rest frame is only a few percent. Encouragingly, however, the points for all the objects fall on the best-fit line to within the combined uncertainties in V^0 and v_r.

The last four entries in Table 1.16, from Lauer et al. (1998), are derived from SBF measurement of the central cD galaxies in the Abell clusters listed, and not from directly resolved globular cluster populations. I have put these in primarily as a consistency check of the SBF analysis technique (Lauer et al. assume a constant value for the GCLF dispersion, and then derive a value of the turnover magnitude which provides the best-fit model for the fluctuation amplitude).

The straight average of the datapoints for the first five entries in the table (the ones with resolved GCLF turnovers) gives $\langle \log\, v_r - 0.2V^0 \rangle = -1.664 \pm 0.018$. Putting in $M_V^0 = -7.33 \pm 0.04$ from Table 1.14, we obtain $H_0 = (74 \pm 4)$ km s^{-1} Mpc^{-1}. The quoted error of course represents only the internal uncertainty of the best-fit line. The true uncertainty is dominated by the absolute uncertainty in the fundamental distance scale (Section 2), which we can estimate (perhaps pessimistically) as ± 0.2 mag once we add all the factors in the chain from parallaxes through the Milky Way to the Virgo/Fornax calibrating region. (For comparison, the scatter of the points about the mean line in Fig. 1.51 is ± 0.25 mag.) A $\pm 0.2-$mag error in M_V^0 translates into $\Delta H_0 = \pm 7$. Thus our end result for H_0 is

$$H_0 \;=\; (74 \pm 4[\mathrm{int}], \pm 7[\mathrm{ext}])\,\mathrm{km\ s^{-1} Mpc^{-1}}\,. \tag{1.55}$$

Taking the mean of all 9 points in the table, including the turnovers deduced from the SBF analysis, would have yielded $H_0 = 72$.

Table 1.16. GCLF Turnover Levels in Remote Galaxies

Cluster	Galaxy	v_r(CMB) (km s^{-1})	V^0	Sources
Virgo	6 gE's	1300	23.73 ± 0.03	1
Fornax	6 gE's	1400	23.85 ± 0.04	1
NGC 5846	NGC 5846	2300	25.08 ± 0.10	2
Coma	IC 4051	7100	27.75 ± 0.20	3,4
Coma	NGC 4874	7100	27.82 ± 0.12	5
Coma	NGC 4881	7100	> 27.6	6
A 262	NGC 705	4650	26.95 ± 0.3	7
A 3560	NGC 5193	4020	26.12 ± 0.3	7
A 3565	IC 4296	4110	26.82 ± 0.3	7
A 3742	NGC 7014	4680	26.87 ± 0.3	7

Sources: (1) This section (2) Forbes et al. 1996a (3) Baum et al. 1997 (4) Woodworth & Harris 1999 (5) Kavelaars et al. 1999 (6) Baum et al. 1995 (7) Lauer et al. 1998

What are the ultimate limits of this distance scale technique? With the HST cameras, many other gE galaxies (BCGs in a variety of Abell clusters) can be added to the graph in Fig. 1.51 out to a limit which probably approaches $cz \sim 10,000$ km s^{-1}. With two or three times as many points, the random uncertainty of the fitted line zeropoint can then be reduced to ± 2 km s^{-1} Mpc^{-1}. Similarly, if the true distance uncertainty to the Fornax and Virgo calibrators can be reduced eventually to ± 0.1 mag, then the total error (internal + external) in H_0 will be reduced to about 7%, making it competitive with any of the other methods in the literature. A more detailed discussion of the uncertainties is given by Whitmore (1997).

This approach to measuring H_0 with the GCLF is the most defensible one on astrophysical grounds. We are deliberately comparing galaxies of strictly similar types (giant ellipticals) over a range of distances, so that we can plausibly argue that the intrinsic differences in their GCLFs, due to any differences in the globular cluster system formation or evolution, will be minimized. Nevertheless, if we wish to be a bit more audacious, we can take a bigger leap of faith by pinning our assumed turnover luminosity M_V^0 *to the Milky Way alone*, arguing that there is no compelling evidence as yet that M_V^0(gE) is systematically different from M_V^0(spiral) (Table 1.14). This assumption would allow us to go directly from our own Galaxy to the Hubble constant in a single leap, bypassing any of the other steps through the Local

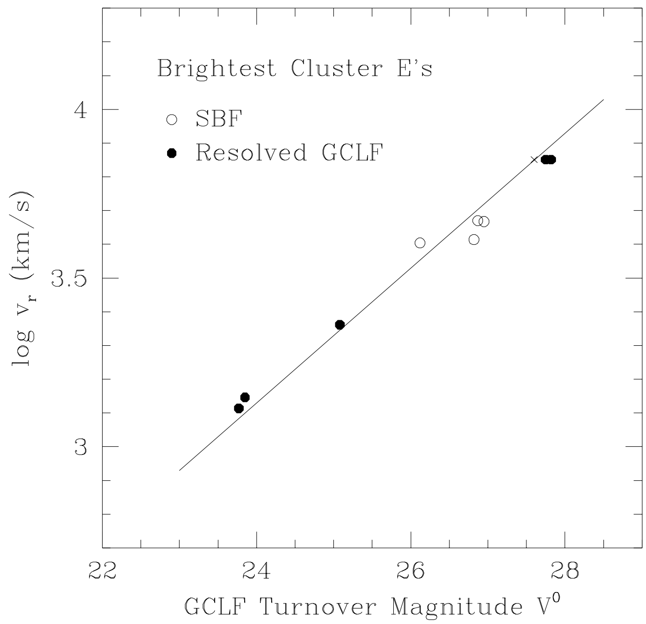

Fig. 1.51. Hubble diagram for globular cluster luminosity functions. The cosmological recession velocity v_r is plotted against the apparent magnitude of the GCLF turnover, for 10 brightest cluster galaxies or groups of galaxies. Solid dots are ones in which the GCLF has been directly resolved down to the turnover point. Open dots are ones in which the turnover level has been deduced by a fit to the surface brightness fluctuation function; see text. The cross is the lower limit for the Coma elliptical NGC 4881. The best-fit straight line (with equal weights to all the solid dots) yields a Hubble constant $H_0 = 74 \pm 8$

Group, Virgo, or Fornax. If we do this with the two Coma ellipticals, using the relevant numbers listed above we obtain $H_0 \simeq 56 - 65$ depending on whether we adopt $M_V^0 = -7.4$ from the entire Milky Way sample or -7.68 from the $r_p > 3$ kpc projected halo sample (Section 5). It is not clear which we should do. In addition, the internal errors are significantly larger than before since only a single galaxy with a rather small GCS population is being used to calibrate the luminosity. However, this Milky Way route should be considered only as an interesting numerical exercise: there is no believable "principle of universality" for GCLFs that we can invoke here, and the true systematic differences between ellipticals and spirals are quantities which must be worked out on observational grounds.

The way that the Hubble constant affects various cosmological parameters is well known and will not be reviewed here (see the textbook of Peebles 1993 or the review of Carroll & Press 1992). For $H_0 \simeq 70$, the Hubble expansion time is $H_0^{-1} = 14.0$ Gyr. If the total mass density has its closure value of $\Omega = 1$, then the true age of the universe is $\tau = (2/3)H_0^{-1} = 9.3$ Gyr, which falls short of the currently calibrated maximum ages of the oldest stars by 3 to 5 Gyr. However, there are strong experimental indications that the overall mass density (dark or otherwise) is only $\Omega_M \simeq 0.1 - 0.3$, such as from the virial masses of rich clusters of galaxies at large radius (Carlberg et al. 1996), the abundances of the light elements (e.g., Mathews et al. 1996), the number density evolution of rich clusters of galaxies (e.g., Bahcall et al. 1997), or the power spectrum of the cosmic microwave background (e.g., Lineweaver et al. 1998).

If $\Lambda = 0$ (no vacuum energy density term) and there are no other terms to add to Ω(global) (Carroll & Press 1992), then the true age of the universe for $\Omega \sim 0.2$ would be $\tau \simeq 13$ Gyr. A value in that range is in reasonable agreement with contemporary estimates of the ages of the oldest stars in the galaxy, measured either by globular cluster ages from isochrone fitting (e.g., VandenBerg et al. 1996; Chaboyer et al. 1998; Carretta et al. 1999), or by thorium radioactive-decay age dating of metal-poor halo stars (Cowan et al. 1997). However, early results from the Hubble diagram analysis of distant supernovae favor a nonzero Ω_Λ and a combined sum $(\Omega_M + \Omega_\Lambda) \sim 1$ (Perlmutter et al. 1997; Riess et al. 1998), though on strictly observational grounds the case is still open. Should we take the somewhat cynical view that those who are hunting for large Ω_Λ are (to quote Erasmus) "looking in utter darkness for that which has no existence"? That would be premature. Many possibilities still exist for additional contributions to Ω, modified inflation models, and so on. The debate is being pursued on many fronts and is certain to continue energetically.

1.8 GLOBULAR CLUSTER FORMATION: IN SITU MODELS

If we knew *what we were doing, it wouldn't be research.*

Anonymous

Understanding how globular clusters form – apparently in similar ways in an amazingly large variety of parent galaxies – is a challenging and long-standing problem. Though it still does not have a fully fleshed-out solution, remarkable progress has been made in the last decade toward understanding the times and places of cluster formation.

The scope of this problem lies in the middle ground between galaxy formation and star formation, and it is becoming increasingly clear that we will need elements of both these upper and lower scales for the complete story to emerge. At the protogalactic (~ 100 kpc) scale, the key question appears to be: How is the protogalactic gas organized? Assuming it is clumpy, what is the characteristic mass scale and mass spectrum of the clumps? Then, at the next level down ($\sim$ 1 kpc), we need to ask how protoclusters form within a single one of these gas clouds. Finally, at the smallest scales ($\lesssim 0.1$ pc), we ask how the gas within protoclusters turns itself into stars. At *each* level it is certain that the answers will involve complex gas dynamics, and full numerical simulations covering the entire $\gtrsim 10^{10}$ dynamic range in mass and length with equal and simultaneous precision are still formidable tasks. We can, however, hope to explore some partial answers. In these next two sections, we will discuss some of the current ideas for massive star cluster formation and ask how successful they are at matching the observations we have now accumulated.

1.8.1 Summarizing the Essential Data

The first theoretical ideas directly relevant to globular cluster formation (in the literature before about 1992) were usually based on the concept that GC formation was in some way a "special" Jeans-mass type of process that belonged to the pre-galactic era (e.g., most notably Peebles & Dicke 1968; Fall & Rees 1985). Such approaches were very strongly driven by the characteristics of the globular clusters in the Milky Way alone, which as we have seen are massive, old, (mostly) metal-poor, and scattered through the halo. Ashman & Zepf (1998) provide an excellent overview of these early models.

All these early models run into severe difficulties when confronted with the rich range of GCS properties in other galaxies, along with the visible evidence of newly formed globular-like clusters in starburst galaxies (Chp. 9 below). For example, traditional models which assumed that globular clusters formed out of *low-metallicity* gas must now be put aside; the plain observational fact is that many or most of the globular clusters in giant E galaxies – and many in large spirals – have healthy metallicities extending up to solar abundance and perhaps even higher. Similarly, no theory can insist that globular clusters are all "primordial" objects in the sense that they formed *only* in the early universe; a wealth of new observations of colliding and starburst galaxies give compelling evidence that $\sim 10^5 - 10^6 M_\odot$ clusters can form in today's universe under the right conditions.

This remarkable new body of evidence has dramatically changed our thinking about cluster formation. It is hard to avoid the view that globular cluster formation is not particularly special, and is in fact linked to the more general process of star cluster formation at any mass or metallicity (e.g. Harris 1996b). Let us summarize the key observational constraints:

- *The Luminosity Distribution Function (LDF):* The number of clusters per unit mass (or luminosity) is rather well approximated by a simple empirical power law $dN/dL \sim L^{-1.8\pm0.2}$ for $L \gtrsim 10^5 L_\odot$. As far as we can tell, this LDF shape is remarkably independent of cluster metallicity, galactocentric distance, parent galaxy type, or other factors such as environment or specific frequency. For smaller masses ($M \lesssim 10^5 M_\odot$), dN/dL

becomes more nearly constant with L, with a fairly sharp changeover at $10^5 L_{\odot}$ (the turnover point of the GCLF). As we saw in the previous Section, the GCLF turnover is similar enough from place to place that it turns out to be an entirely respectable standard candle for estimating H_0.

- *The Metallicity Distribution Function:* The number of clusters at a given metallicity differs significantly from one galaxy to another. In the smallest dwarf galaxies, a simple, single-burst model leaving a low metallicity population gives a useful first approximation. In large spiral galaxies and in many large ellipticals, clearly bimodal MDFs are present, signalling at least a two-stage (or perhaps multi-stage) formation history. And in some giant ellipticals such as NGC 3311 (Secker et al. 1995) or IC 4051 (Woodworth & Harris 1999), the MDF is strongly weighted to the high-[Fe/H] end, with the metal-poor component almost completely lacking. What sequence of star formation histories has generated this variety?
- *Specific Frequencies:* The classic "S_N problem" is simply stated: Why does the relative number of clusters differ by more than an order of magnitude among otherwise-similar galaxies (particularly elliptical galaxies)? Or is there a hidden parameter which, when included, would make the true cluster formation efficiency a more nearly universal ratio?
- *Continuity of cluster parameters:* Aside from the points mentioned above, one obvious observational statement we can now make (Harris 1996b) is that in the 3-space of cluster mass, age, and metallicity (M, τ, Z), we can find star clusters within *some* galaxy with almost every possible combination of those parameters. The Milky Way is only one of the diverse cluster-forming environments we can choose to look at. In the physical properties of the clusters themselves, there are no sudden transitions and no rigid boundaries in this parameter space.

1.8.2 The Host Environments for Protoclusters

What framework can we assemble to take in all of these constraints? Having been forced to abandon the view that globular cluster formation is a special, early process, let us make a fresh start by taking the opposite extreme as a guiding precept:

All types of star clusters are fundamentally similar in origin, and we will not invoke different formation processes on the basis of mass, age, or metallicity.

The immediate implication of this viewpoint is that we should be able to learn about the formation of globular clusters by looking at the way star clusters are forming today, both in the Milky Way and elsewhere. This same point was argued on an empirical basis in a series of papers by Larson (e.g., 1988, 1990a,b, 1993, 1996) and has now turned into the beginnings of a more quantitative model by Harris & Pudritz (1994) and McLaughlin & Pudritz (1996); see also Elmegreen & Falgarone (1996) and Elmegreen &

Efremov (1997) for an approach which differs in detail but starts with the same basic viewpoint. In these papers, we can find the salient features of cluster formation which are relevant to this new basis for formation modelling:

- Star clusters are seen to form out of the very densest clumps of gas within giant molecular clouds (GMCs).
- In general, the mass contained within any one protocluster is a small fraction (typically 10^{-3}) of the total mass of its host GMC. The formation of bound star clusters, in other words, is an unusual mode which seems to require a *large surrounding reservoir of gas.*
- Many or most *field stars* within the GMC are also expected to form within small groups and associations, as recent high-resolution imaging studies of nearby star forming regions suggest (e.g., Zinnecker et al. 1993; Elmegreen et al. 1999). Most of these clumps are likely to become quickly unbound (within a few Myr) after the stars form, presumably because much less than 50% of the gas within the clump was converted to stars before the stellar winds, ultraviolet radiation, and supernova shells generated by the young stars drive the remaining gas away. Observationally, we see that typically *within one GMC* only a handful ($\sim 1 - 10$) of protoclusters will form within which the star formation efficiency is high enough to permit the cluster to remain gravitationally bound over the long term. This empirical argument leads us to conclude that on average, perhaps $\lesssim 1\%$ of the host GMC mass ends up converted into bound star clusters; a much higher fraction goes into what we can call "distributed" or field-star formation.[10]
- The larger the GMC, the more massive the typical star cluster we find in it. In the Orion GMC, star clusters containing $10^2 - 10^3 M_{\odot}$ have recently formed, within a GMC of $\sim 10^5 M_{\odot}$. But, for example, in the much more massive 30 Doradus region of the LMC, a $\gtrsim 2 \times 10^4 M_{\odot}$ cluster (R136) has formed; this young object can justifiably be called a young and more or less average-sized globular cluster. Still further up the mass scale, in merging gas-rich galaxies such as the Antennae (see Chp. 9 below), $\gtrsim 10^5 M_{\odot}$ young star clusters have formed within the $10^7 - 10^8 M_{\odot}$ gas clouds that were accumulated by collisional shocks during the merger.
- A GMC, as a whole, has a clumpy and filamentary structure with many embedded knots of gas and denser gas cores. Its internal pressure is dominated by the energy density from turbulence and weak magnetic field; direct thermal pressure is only a minor contributor. (That is, the internal motions of the gas within the GMC are typically an order of magnitude higher than would be expected from the temperature of the gas alone; other sources of energy are much more important.) Thus, the GMC lifetime as a gaseous entity is at least an order of magnitude longer than

[10] As we will see later, McLaughlin (1999) arrives at a fundamentally similar conversion ratio of ~ 0.0025 by comparing the total mass in globular clusters to total galaxy mass (stars plus gas) for giant E galaxies.

would be expected from radiative cooling alone; the GMC cannot cool and collapse until the internal magnetic field leaks away (e.g., Carlberg & Pudritz 1990; McKee et al. 1993), unless external influences cause it to dissipate or disrupt sooner. The gas within the GMC therefore has plenty of time to circulate, and the dense cores have relatively large amounts of time to grow and eventually form stars.

- The dense gas cores within GMCs are particularly interesting for our purposes, because they are the candidates for proto star clusters. Their mass spectrum should therefore at least roughly resemble the characteristic power-law mass distribution function that we see for the star clusters themselves. And indeed, they do – perhaps better than we could have expected: the *directly observed* mass distribution functions of the gaseous clumps and cores within GMCs follow $dN/dM \sim M^{-\alpha}$, with mass spectral index α in the range $1.5-2.0$. The same form of the mass distribution function, and with exponent in the same range, is seen for young star clusters in the LMC, the Milky Way, and the interacting galaxies within which massive clusters are now being built (see below). As we have already seen, the luminosity distribution function for globular clusters more massive than $\sim 10^5 M_\odot$ follows the same law, with minor variations from one galaxy to another. On physical grounds, the extremely high star formation efficiency ($\sim 50\%$ or even higher) necessary for the formation of a bound star cluster is the connecting link that guarantees the similarity of the mass distributions – the input mass spectrum of the protoclusters, and the emergent mass spectrum of the young star clusters (see Harris & Pudritz 1994).

The clues listed above provide powerful pointers toward the view that globular clusters formed within GMCs by much the same processes that we see operating today within gas-rich galaxies. The single leap we need to make from present-day GMCs to the formation sites of globular clusters is simply one of mass scale. Protoglobular clusters are necessarily in the range $\sim 10^4 - 10^6 M_\odot$. Then by the scaling ratios mentioned above, they must have formed within very large GMCs – ones containing $\sim 10^7 - 10^9 M_\odot$ of gas and having linear sizes up to ~ 1 kpc (Harris & Pudritz 1994). These postulated "supergiant" molecular clouds or SGMCs are larger than even the most massive GMCs found in the Local Group galaxies today by about one order of magnitude. But in the pregalactic era, they must have existed in substantial numbers within the potential wells of the large protogalaxies, as well as being scattered in sparser numbers between galaxies.

The SGMCs, in size and mass, obviously resemble the pregalactic 'fragments' invoked two decades ago by Searle (1977) and Searle & Zinn (1978). Their reason for doing so was driven by the need for appropriate environments in which place-to-place differences in local chemical enrichment could arise, thus producing a globular cluster system with a large internal scatter in metallicity and little or no radial gradient. These same dwarf-galaxy-sized gas clouds also turn out to be just what we need to produce star clusters with the right mass scale and mass spectrum. Whether we call them SGMCs, protogalactic subsystems, or pregalactic fragments, is a matter only of terminology (Harris 1996b).

We might wish to claim that globular cluster formation does preferentially belong to a "special" epoch – the early universe of protogalaxies. The grounds for this claim are simply that this was the epoch when by far the most gas was available for star formation, and SGMCs could be assembled in the largest numbers. Many Gigayears later, in today's relatively star-rich and gas-poor universe, most of the gas is in the form of (a) rather small GMCs (within spiral and irregular galaxies), which can produce only small star clusters, (b) the much lower-density ISM within the same galaxies, and (c) hot X-ray halo gas in giant ellipticals and rich clusters of galaxies, within which star formation cannot take place. Globular-sized cluster formation can still happen, but only in the rare situations where sufficiently large amounts of relatively cool gas can be assembled.

In short, the populations of globular clusters in galactic halos can be viewed as byproducts of the star formation that went on in their highly clumpy protogalaxies. Direct observations of high-redshift galaxies confirm the basic view that large galaxies form from hierarchical merging of smaller units (e.g., Pascarelle et al. 1996; Madau et al. 1996; Steidel et al. 1996; van den Bergh et al. 1996; Glazebrook et al. 1998; and references cited there). Even the smaller systems at high redshift appear to be undergoing star formation both before and during their agglomerations into larger systems. One can scarcely improve on Toomre's (1977) prescient remark that there was almost certainly "a great deal of merging of sizeable bits and pieces (including many lesser galaxies) early in the career of every major galaxy".

It is also apparent that, if any one of these SGMCs were to avoid amalgamation into a larger system and were left free to evolve on its own, it would end up as a normal dwarf galaxy – either spheroidal or irregular, depending on what happens later to its gas supply. The identification of the dwarf ellipticals that we see today as leftover "unused" pieces, some of them with globular clusters of their own, is also a natural step (Zinn 1980, 1993b; Mateo 1996). However, it seems considerably riskier to assume further that the halo of the Milky Way, or other large galaxies, was simply built by the accretion of dwarfs that had *already* formed most of their stars (e.g. Mateo 1996). Much of the merging and amalgamation of these building blocks must have happened early enough that they were still mostly gaseous – as indeed, some still are today (see below).

1.8.3 A Growth Model for Protoclusters

If we have convinced ourselves that globular clusters need large, local reservoirs of gas within which to form, how does the host SGMC actually convert a small portion of its gas into protoclusters? A full understanding of the process must surely plunge us deeply into the complex business of gas magnetohydrodynamics. Yet at its basis, the driving mechanism can reasonably be expected to be a simple one. The most successful, and most quantitative, approach we have at present is the model of McLaughlin & Pudritz (1996) based on the precepts in Harris & Pudritz (1994). The basis of this model is that the dense clumps of gas circulating within the GMC will build up into protoclusters by successive collision and agglomeration. Collisional growth is a well understood process which arises in many comparable situations (such as planetesimal growth in the protosolar nebula, or the buildup of a cD galaxy from its smaller neighbors). It is, in addition, important to note that the clumpy, filamentary structure of the typical GMC and its high internal motions will guarantee that collisional agglomeration will be taking place regardless of whatever else is happening within the cloud. Furthermore, the long lifetime of the GMC against cooling and collapse (see above) suggests that the growth process will have a sensibly long time to work.

We now briefly outline the essential steps in the collisional growth process, as developed especially by Field & Saslaw (1965), Kwan (1979), and McLaughlin & Pudritz (1996) for protoclusters. It is schematically outlined in Fig. 1.52. The GMC is idealized as containing a large supply of small gas particles of mass m_0 (small dots) which circulate within the cloud. As a very rough estimate of their mass range, we might perhaps think of the m_0's as physically resembling the $\sim 100 M_\odot$ dense cores found in the Milky Way GMCs. Whenever two clouds collide, they stick together and build up larger clumps. Eventually, when a clump gets large enough, it terminates its growth by going into star formation and turning into a star cluster; the unused gas from the protocluster will be ejected back into the surrounding GMC, thus partly repopulating the supply of m_0's. At all times, the total mass in the protoclusters is assumed to be much less ($\lesssim 1\%$) than the total GMC mass, so the supply of m_0's is always large.

The growth of the protoclusters can then be followed as the sum of gains and losses, through a rate equation which generates a clump mass spectrum dn/dm. The number at a given mass m *decreases* whenever (a) a cloud at m combines with another at m' to form a bigger one; or (b) a cloud at m turns into stars, at a rate determined by the cooling timescale (denoted τ_m). Conversely, the number at m *increases* whenever (c) two smaller clouds m', m'' combine to form one cloud at m, or (d) larger clouds disrupt to form smaller ones, according to a "replenishment" spectrum $r(m)$. The sum of all four processes operating together as time goes on creates the output mass spectrum. Small clouds always vastly outnumber the larger ones, so that in a statistical sense, the larger ones almost always grow by absorbing much smaller clouds. By contrast, mutual collisions between two already-massive clouds are relatively rare.

The basic theory of Field & Saslaw (1965), which assumes an initial population of identical m_0's and velocities, and simple geometric collisions with

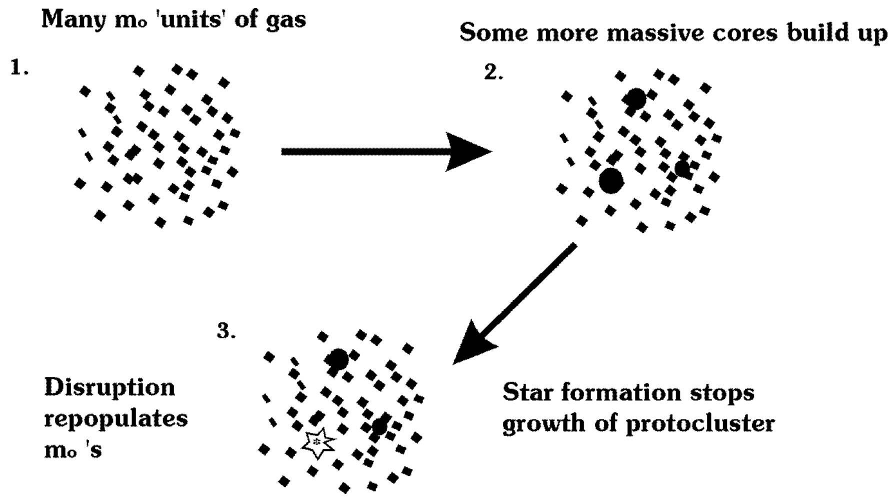

Fig. 1.52. Schematic illustration of the growth of protoclusters within a GMC by collisional agglomeration

no disruption or cooling, yields a characteristic distribution $dn/dm \sim m^{-1.5}$. Kwan's (1979) development of the model shows that a range of mass exponents (mostly in the range $\sim 1.5 - 2.0$) can result depending on the way the internal cloud structure and cloud velocity distribution vary with m. McLaughlin & Pudritz further show the results of including the star formation timescale and disruption processes. The detailed shape of the emergent mass spectrum is controlled by two key input parameters:

(a) The first parameter is the dependence of cloud lifetime on mass, which is modelled in this simple theory as $\tau_m \sim m^c$ for some constant exponent $c < 0$. In this picture, more massive clouds – or at least the dense protocluster regions within them – should have equilibrium structures with shorter dynamical timescales and shorter expected lifetimes before beginning star formation, thus $c < 0$. Cloud growth in this scenario can be thought of as a stochastic race against time: large clouds continue to grow by absorbing smaller ones, but as they do, it becomes more and more improbable that they can continue to survive before turning into stars. The consequence is that at the high-mass end, the slope of the mass spectrum dn/dm gradually steepens.

(b) The second parameter is the cloud lifetime τ against star formation divided by the typical cloud-cloud collision time, which is $\tau_0 \simeq m_0/(\rho\sigma_0 v_0)$. Here ρ is the average mass density of the clouds, σ_0 is the collision cross section of two clouds at m_0, and v_0 is the typical relative velocity between clouds. Denote $\tau_\star$ as a fiducial cloud lifetime, and then define the timescale ratio $\beta = \tau_\star/\tau_0$. For the collisional growth model to give the result that we need, we must have $\beta \gg 1$. That is, the internal timescale of the cloud governing how soon it can cool, dissipate, and go into star formation, must be significantly longer than the cloud-cloud collision time. If it is not (i.e. suppose $\beta \sim 1$), it would mean that the protocluster clouds would turn into stars roughly as fast as they could grow by collision, and thus the emergent star clusters would all have masses not much larger than m_0 itself. The larger the value of β, the shallower the slope of the spectrum dn/dm will be, and the further up to higher mass it will extend (though at high mass, it will get truncated by the decrease of τ_m, as noted above).

In summary, the timescale ratio β influences the basic slope of the power-law mass spectrum, while the exponent c determines the upper-end falloff of the spectrum slope and thus the upper mass limit of the distribution. Changes in other features of the model, such as the initial mass distribution or the details of the replenishment spectrum $r(m)$, turn out to have much less important effects. (For example, rather than assuming all the clouds to have the same mass m_0, one could assume some initial range in masses. However, the main part of the emergent mass spectrum that we are interested in is where m is orders of magnitude larger than m_0, where the memory of the initial state has been thoroughly erased by the large number of collisions. See McLaughlin & Pudritz for additional and much more detailed discussion.)

Fits of this model to the observed LDFs of the well observed globular cluster systems in the Milky Way, M31, and M87 show remarkably good agreement for the clusters *above the turnover point* (see Figs. 1.53 and 1.54). The exponent c is $\simeq -0.5$ for all of them, while the ratio β is $\simeq 115$ for M87 but $\simeq 35$ for the steeper LDFs in the two spiral galaxies.

Notably, this simple theory reproduces the upper $\sim 90\%$ of the cluster mass distribution extremely well, but it does not reproduce the abrupt flattening of the LDF at the low-mass end. A natural suspicion is, of course, that the observed distribution contains the combined effects of more than 10^{10} years of dynamical erosion on these clusters in the tidal field of the parent galaxy, which would preferentially remove the lower-mass ones. In other words, the formation model *must* predict "too many" low-mass clusters compared with the numbers we see today. A valuable test of this idea would therefore be to compare the model with an LDF for a much younger set of clusters which can plausibly be assumed to have a small dispersion in age and which have had much less time to be damaged by dynamical evolution.

The best available such data at present are for the newly formed star clusters in recent mergers such as NGC 4038/4039 (Whitmore & Schweizer 1995) and NGC 7252 (Miller et al. 1997). An illustrative fit of the collisional-growth theory to these LDFs is shown in Fig. 1.55. Here, the low-mass end of the cluster mass distribution is more obviously present in significant numbers, and the overall distribution provides a closer global match to the theory.

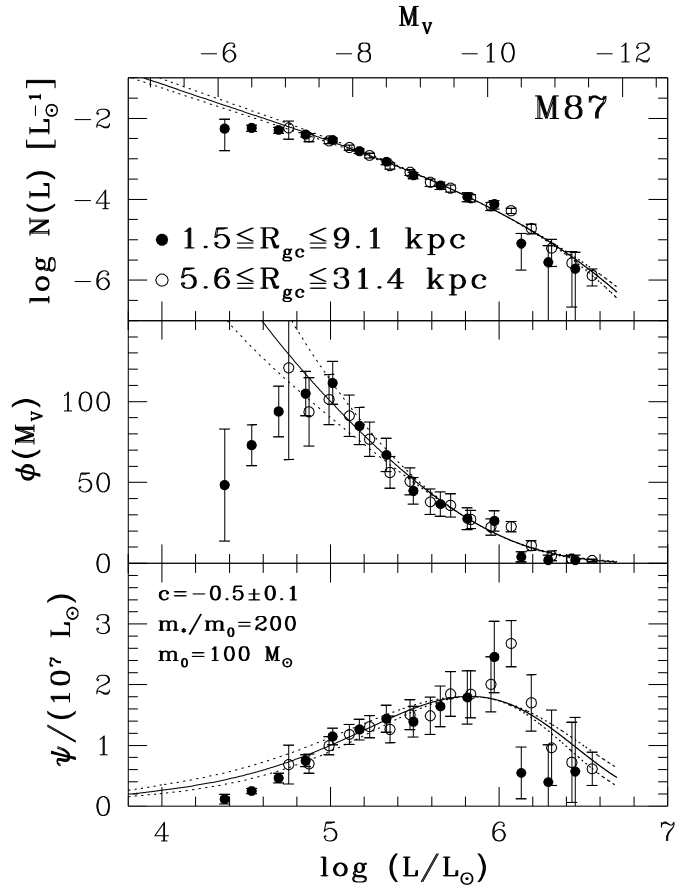

Fig. 1.53. Fit of the collisional-growth model of the mass distribution function to the LDF for M87 (McLaughlin 1999, private communication). The *top panel* shows the LDF itself, for a collisional growth model with "particle size" $m_0 = 100 M_\odot$ and timescale exponent $c = -1/2$ (see text). Solid dots show the observed LDF for globular clusters in the inner halo of M87, open circles for the outer halo. The *middle panel* shows the luminosity distribution in its more traditional form as the number per unit magnitude (GCLF); and the *bottom panel* shows the luminosity-weighted GCLF (essentially, the amount of integrated light contained by all the clusters in each bin). Note that the model line in each case fits the data well for log $(L/L_\odot) \gtrsim 4.7$ (the upper 90% of the mass range) but predicts too many clusters at fainter levels

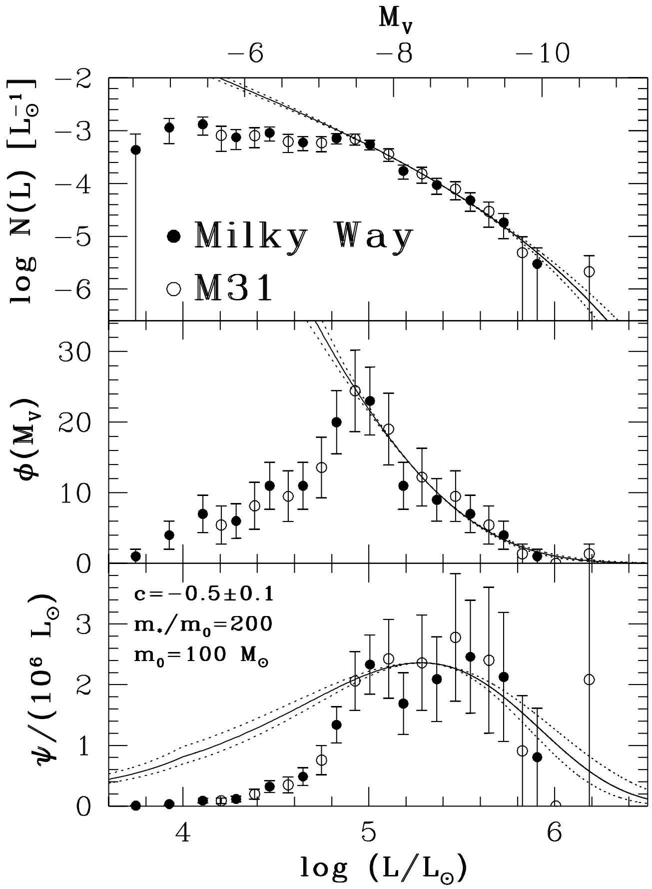

Fig. 1.54. Fit of the collisional-growth model of the mass distribution function to the LDFs for M31 and the Milky Way (McLaughlin 1999, private communication). Panels are the same as in the previous figure

Encouragingly, the gradual steepening of the LDF toward the high-mass end is present here as well, just as we expect from the model.

The fit shown in Fig. 1.55 is deceptively good, however, because it assumes that the clusters shown have a *single age*, i.e. it assumes all of them were formed in one short-duration burst. In this view, any differences in the measured colors of the clusters are ascribed as due to internal random differences in reddening rather than age. For a relatively young merger like NGC 4038/39, the clusters must surely have differences in both reddening and age, but clearly separating out these two factors is difficult. If the clusters formed in one burst, the few-Myr age differences would generate unimportant scatter in the LDF. At the opposite extreme, the ages could range over the entire ~ 200 Myr duration of the merger event. From the $(U-V, V-I)$ color dis-

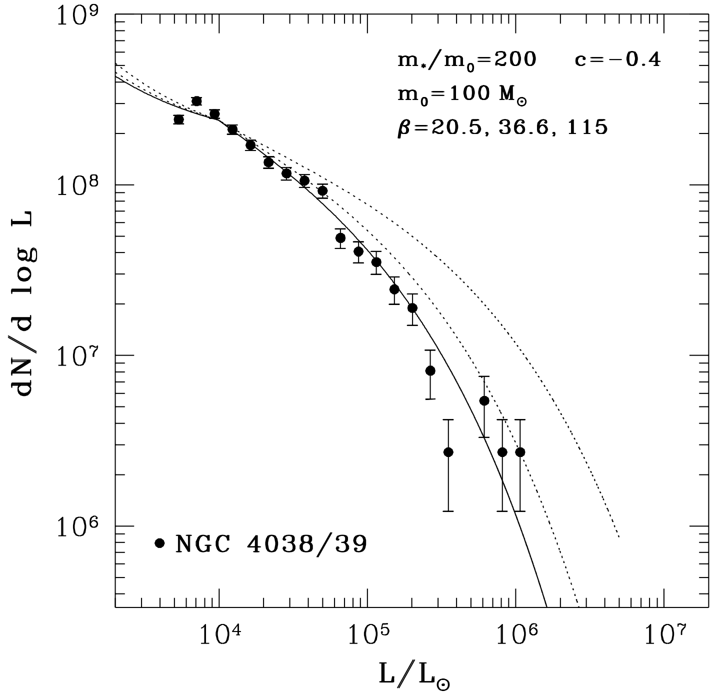

Fig. 1.55. Fit of the collisional-growth model of the mass distribution function to the young star clusters in the merging galaxies NGC 4038/39 (McLaughlin 1999, private communication). The curves running from left to right correspond to the three $\beta-$values listed

tribution and the presence of HII regions around many clusters, Whitmore & Schweizer (1995) argue that their age range is probably 3 to 30 Myr if they have solar metallicity, and that reddening differences probably make the estimated age range look artificially broad. Meurer (1995) provides a brief but perceptive discussion of the effect of any age range on the LDF, showing that its shape can be systematically distorted by the different rates at which younger and older clusters will fade over the same length of time.

A more extensive analysis is presented by Fritze-von Alvensleben (1998, 1999), who uses a different set of cluster models than Whitmore & Schweizer, and a different assumed metallicity of $0.5Z_{\odot}$. She deduces from the mean $(V-I)$ colour of the sample a mean cluster age of 200 Myr, similar to the estimated time of the last pericenter passage of the two galaxies (Barnes 1988). Notably, she also finds that by keeping only the most compact objects (effective radii $R_{eff} < 10$ pc), the resulting LDF is then proportionately less populated at the *faint* end, curving over more strongly than would be expected from the model of Fig. 1.55. This effect is enhanced even further if the ages of the clusters are *individually* estimated from their $(V-I)$ colors (assuming, perhaps wrongly, that all of them have the same reddening). If all the clusters are individually age-faded to 12 Gyr, the resulting GCLF strongly resembles the classic Gaussian in number per unit magnitude, with the expected turnover at $M_V \sim -7$.

Fritze-von Alvensleben's analysis provides interesting evidence that the GCLF of globular clusters may take on its "standard" Gaussian-like form in number per unit magnitude at a very early stage. We can speculate that most of the faint clusters that should theoretically form in large numbers will appear in diffuse clumps that dissolve quickly away into the field during the first few $\sim 10^8$ y, leaving the low-mass end of the distribution depleted as the observations in all old galaxies demand.

Much remains to be investigated in more detail. For example, in the formation model outlined above, the key quantities (c, β) are free parameters to be determined by the data; more satisfactorily, we would like to understand their numerical ranges from first principles more accurately than in the present rough terms. In addition, it is not clear what determines the $\gtrsim 10^3$ ratio of host GMC mass to typical embedded cluster mass, or what this ratio might depend on. Nevertheless, this basic line of investigation appears to be extremely promising.

1.8.4 The Specific Frequency Problem: Cluster Formation Efficiency

Another of the outstanding and longest-standing puzzles in GCS research, as described above, is the S_N problem: in brief, why does this simple parameter differ so strongly from place to place in otherwise similar galaxies?

Let us first gain an idea of what a "normal" specific frequency means in terms of the mass fraction of the galaxy residing in its clusters. Define an efficiency parameter e as number of clusters per unit mass,

$$e = \frac{N_{cl}}{M_g} \tag{1.56}$$

where N_{cl} is the number of clusters and M_g is the total gas mass in the protogalaxy that ended up converted into stars (that is, into the visible light). Then we have $S_N = \text{const} \times e$ or

$$e = \frac{S_N}{8.55 \times 10^7\,(M/L)_V} \tag{1.57}$$

where $(M/L)_V \sim 8$ is the visual mass-to-light ratio for the typical old-halo stellar population. Most of the reliable S_N measurements are for E galaxies, and for a baseline average $S_N^0 \simeq 3.5$ (Section 6), we immediately obtain a fiducial efficiency ratio $e_0 \simeq 5.1 \times 10^{-9} M_\odot^{-1}$, or $e_0^{-1} \sim 2 \times 10^8 M_\odot$ per cluster. For an average cluster mass $\langle M_{cl} \rangle = 3 \times 10^5 M_\odot$ (from the Milky Way sample), the typical mass fraction in globular clusters is $e_0 \langle M_{cl} \rangle = 0.0015$, or 0.15%. Although this ratio is encouragingly close to what we argued empirically for GMCs in the previous section, we would have to allow for galaxy-to-galaxy differences of factors of 5 or so both above and below this mean value, in order to accommodate all the E galaxies we know about. Invoking simple differences in the efficiency with which gas was converted into star clusters remains a possibility, but is an uncomfortably arbitrary route.

The alternative possibility is to assume that the initial cluster formation efficiency was more or less *the same in all environments*, but that the higher$-S_N$ galaxies like M87 did not use up all their initial gas supply (in a sense, we should view such galaxies not as "cluster-rich" but instead as 'field-star poor"). This view has been raised by Blakeslee (1997), Blakeslee et al. (1997), Harris et al. (1998a), and Kavelaars (1999), and is developed in an extensive analysis by McLaughlin (1999). Simply stated, this view requires that the globular clusters formed in numbers that were in direct proportion to the *total available gas supply* within the whole protogalaxy, and *not* in proportion to the amount of gas that actually ended up in stars of all types.

McLaughlin (1999) defines a new efficiency parameter as a ratio of masses as follows:

$$\epsilon = \frac{M_{cl}}{M_\star + M_{gas}} \tag{1.58}$$

where $M_\star$ is the mass now in visible stars, while M_{gas} is the remaining mass in or around the galactic halo. This residual gas was, by hypothesis, originally part of the protogalaxy. In most large galaxies, $M_{gas} \ll M_\star$. However, for the giant E galaxies with large amounts of hot X-ray halo gas, the additional M_{gas} factor can be quite significant, especially for cD galaxies and other BCGs (brightest-cluster galaxies). Under this hypothesis, the observed specific frequency represents the *proportion of unused or lost initial gas mass* (Harris et al. 1998a):

$$f_M(\text{lost}) = \frac{M_{gas}}{M_\star + M_{gas}} = \left(1 - \frac{S_N^0}{S_N}\right) . \tag{1.59}$$

If, for example, we adopt a baseline $S_N^0 = 3.5$, then a high$-S_N$ BCG like M87 would have $f_M \sim 0.7$, implying that a startlingly high amount – almost three-quarters – of its initial protogalactic mass went unconverted during star formation. Blakeslee (1997) hypothesizes that much of the gas in the original distribution of pregalactic clouds may have been stripped away to join the general potential well of the galaxy cluster during the violent virialization stage. Harris et al. (1998a) suggest instead that a large amount of gas within the proto-BCG may have been expelled outward in a galactic wind during the first major, violent burst of star formation. Both of these mechanisms may have been important, particularly in rich clusters of galaxies. In either case, this ejected or stripped gas would now occupy the halo and intracluster medium in the form of the well known hot gas detectable in X-rays.

By a detailed analysis of three giant E galaxies with high-quality surface photometry and data for X-ray gas and globular cluster populations (NGC 4472 and M87 in Virgo, and NGC 1399 in Fornax), McLaughlin (1999) finds that the mass ratio ϵ is much the same in all three, at $\langle\epsilon\rangle = 0.0025 \pm 0.007$. This result also turns out to hold at the *local* as well as the *global* level, at any one radius of the halo as well as averaged over the whole galaxy. M87, with the highest specific frequency, also has the most halo gas; its proportions of stellar and gas mass are to first order similar. For most galaxies (non-BCGs), the halo gas makes only a minor contribution and S_N is a more nearly correct representation of the mass ratio ϵ.

M87 is only one of many BCGs, and these galaxies as a class are the ones with systematically high S_N and high X-ray luminosity. The direct observations of their cluster populations (see Harris et al. 1998a for a summary of the data) show that the total cluster population scales with visual galaxy luminosity as $N_{cl} \sim L_V^{1.8}$; while the X-ray luminosity scales as $L_X \sim L_V^{2.5\pm0.5}$. With the *very* crude (and, in fact, incorrect) assumption that the X-ray gas mass scales directly as L_X, we would then expect that the ratio of gas mass to stellar mass increases with galaxy size roughly as $(M_{gas}/M_\star) \sim L_V^{1.5}$, which turns out to match the way in which S_N systematically increases with luminosity for BCGs.

McLaughlin (1999) analyzes these scaling relations in considerably more detail, putting in the fundamental-plane relations for gE galaxies (scale size, internal velocity dispersion, and mass-to-light ratio as functions of L_V) as well as the way that the gas mass scales with X-ray luminosity, temperature, and halo scale size. When these are factored in, he obtains

$$S_N \sim \frac{N_{cl}}{L_{gal}} \sim \epsilon\,(1 + \frac{M_{gas}}{M_\star})L_{gal}^{0.3}\,. \tag{1.60}$$

This correlation is shown as the model line in Fig. 1.56. The last term $L_{gal}^{0.3}$ accounts for the systematic increase in mass-to-light ratio with galaxy luminosity: bigger ellipticals have more mass per unit light and thus generated more *clusters* per unit light under the assumption that ϵ (the cluster mass fraction) was constant. The term in parentheses $(1 + \frac{M_{gas}}{M_\star})$ accounts for the presence of high-temperature gas in the halo. As suggested above, in this first-order picture the gas is assumed to have belonged to the protogalaxy, but was heated at an early stage during star formation (by an energetic galactic wind, or tidal stripping of the SGMCs?), and left to occupy the dark-matter potential well in a shallower distribution than the halo stars.

An intriguing implication – and requirement – of this overall view would be that the main epoch of globular cluster formation must be *early* in the star-forming stage: that is, the clusters form in numbers that are in direct proportion to the total gas supply, and do so ahead of most of the stars. Then, in the most massive protogalaxies, the star formation is interrupted before it can run to completion, leaving behind lots of clusters as well as a considerable amount of hot, diffuse gas surrounding the visible galaxy.

It is probably not unreasonable to suppose that the star formation would proceed soonest and fastest within the densest clumps of gas (i.e., the protoclusters). For example, observational evidence from the color-magnitude diagram for R136 (in the LMC, and certainly the nearest example of a young globular cluster) indicates that it has taken $\lesssim 3$ Myr to form, from the lowest-mass stars to the highest-mass ones (Massey & Hunter 1998). It would be premature to claim that this result would be typical of all massive star clusters, but it favors the view that the densest clusters can form rapidly, over times far shorter than the $\sim 10^8$ y dynamical timescale of the protogalaxy.

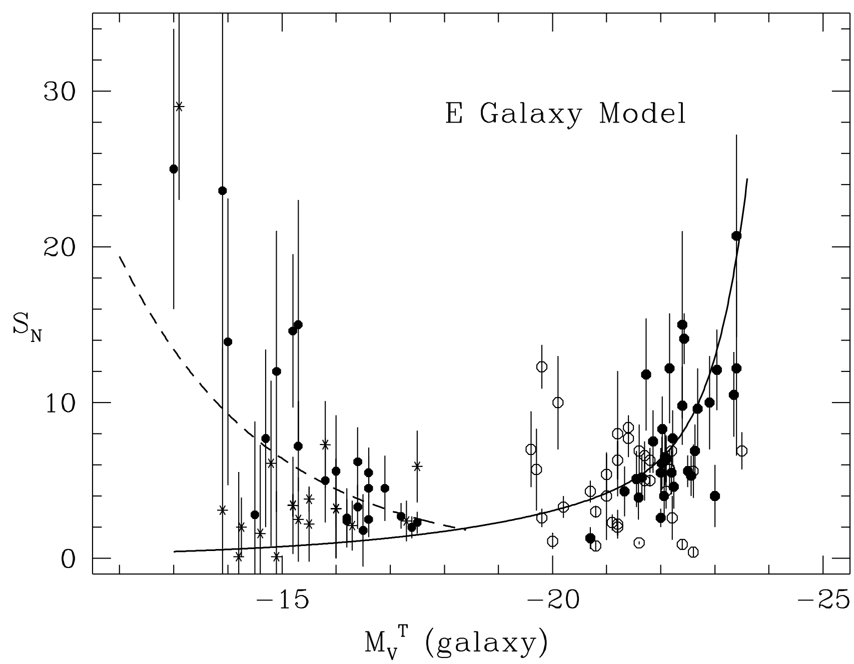

Fig. 1.56. Specific frequency against luminosity for elliptical galaxies. The model line (solid curve, from McLaughlin 1999), assumes $\epsilon = 0.0025 =$ constant, i.e. that globular clusters were formed in these galaxies in direct proportion to the original gas mass of the protogalaxy. BCG galaxies (see text) are the solid symbols at high luminosity, while normal E's are plotted as open symbols. At the low-luminosity end, nucleated dE's are the solid symbols and non-nucleated dE's are the starred symbols. The dashed line assumes the Dekel/Silk model of mass loss, which strongly affects the smaller dwarfs. See text for discussion

1.8.5 Intergalactic Globular Clusters: Fact or Fancy?

White (1987) raised the intriguing idea that the BCGs in rich clusters of galaxies might be surrounded by populations of globular clusters that are not gravitationally bound to the BCG itself, but instead belong to the general potential well of the whole cluster. West et al. (1995) have pursued this concept in detail, suggesting that the huge globular cluster populations around most BCGs might be dominated by such "intergalactic" clusters.

At some level, free-floating intergalactic clusters must be present: strong evidence now exists for intergalactic Population II stars in the Fornax and Virgo clusters (e.g., Theuns & Warren 1997; Ferguson et al. 1998; Ciardullo et al. 1998), and some globular clusters should accompany these stars if they have been tidally stripped from the cluster galaxies. It is only a matter of time before individual cases are detected in nearby clusters of galaxies (accidentally or otherwise). Systematic searches are also underway for "orphan" collections of globular clusters near the centers of clusters of galaxies that do not have central BCGs.

It is, however, still totally unclear whether or not such objects would exist in sufficient numbers to affect the BCGs noticeably. Harris et al. (1998a) use the Virgo giant M87 as a detailed case study to reveal a number of critical problems with this scenario. To *create* the high S_N values seen in the BCGs, the putative intergalactic globulars would have to be present in large numbers *without adding contaminating field-star light of their own*; that is, the intergalactic material would itself need to possess a specific frequency in the range $S_N \sim 100$ or even higher. Furthermore, the "extra" clusters in M87 and the other BCGs are not just distributed in the outermost halo where the larger-scale intergalactic population might be expected to dominate; M87 has more clusters at *all* radii, even in the central few kpc. If all of these are intergalactic, then they would need to be concentrated spatially like the central galaxy, which is contradictory to the original hypothesis.

A promising new way to search for intergalactic material is through radial velocity measurement: such clusters (or planetary nebulae, among the halo stars) would show up as extreme outliers in the velocity histogram. Preliminary velocity surveys of the GCS around the Fornax cD NGC 1399 (Kissler-Patig 1998; Minniti et al. 1998; Kissler-Patig et al. 1999) are suggestive of either an intergalactic component with high velocity dispersion, or possibly contamination from neighboring galaxies. By comparison, few such objects appear in the M87 cluster velocity data (see Cohen & Ryzhov 1997; Harris et al. 1998a). Larger statistical samples of velocities are needed.

1.8.6 The Relevance of Cooling Flows

An idea proposed some time ago by Fabian et al. (1984) was that the large numbers of "extra" clusters in M87 could have condensed out of the cooling flow from the X-ray gas. By inference, other high$-S_N$ galaxies – mainly the central BCGs with large X-ray halos – should be ones with high cooling flows.

This hypothesis has become steadily less plausible. The biggest *a priori* difficulty is that to produce an increased S_N this way, we would have to invoke particularly efficient globular cluster formation (relative to field-star formation) out of the hot, dilute X-ray gas – exactly the type of situation that we would expect should be least likely to do this. However, probably the strongest argument against such a scheme is that there is not the slightest observational evidence that young, massive star clusters exist in any of the pure cooling-flow galaxies (for more extensive discussion, see Harris et al. 1995; Bridges et al. 1996; Holtzman et al. 1996). The true mass dropout rates from these cooling flows remain uncertain, and may in any case be considerably less than was thought in the early days of the subject.

A summary of the observed cases is given in Table 1.17: the Abell cluster designation and central galaxy NGC number (if any) are given in the first two columns (here "BCG" denotes brightest cluster galaxy), the presence or absence of young clusters in the third column, and the deduced cooling flow rate ($M_\odot$ per year) in the last column. GCS data are taken from the three papers cited above, and the cooling flow rates from Allen & Fabian (1997), McNamara & O'Connell (1992), and Stewart et al. (1984). The only two cases which are seen to contain young globular clusters in their central regions (the BCGs in Abell 426 and 1795) are also the two which represent accreting or interacting systems, with large amounts of cooler gas present as well. It seems more probable that it is the cooler, infalling gas that has given rise to the recent starbursts in these giant galaxies, rather than the hot X-ray gas halo.

It is interesting to note that the cooling-flow scenario can, in some sense, be viewed historically as exactly the opposite of the view discussed above involving early mass loss from the BCGs. Rather than suggesting that large number of clusters formed at later times out of the hot X-ray gas, we now suggest that the clusters formed in their large numbers at an early stage, in direct proportion to the original reservoir of gas; but that considerable unused gas was ejected or left out in the halo to form the X-ray halo – which is now the source of the cooling flow.

1.8.7 Dwarf Ellipticals

The dwarf elliptical galaxies present a special puzzle. We see that extreme S_N values are also found among the lowest-luminosity dwarf ellipticals, at the opposite end of the galaxy size scale; but ones with low S_N are present too (Section 6). It seems likely that *early mass loss* is responsible for the high S_N values in these dwarfs. As discussed in Durrell et al. (1996a) and McLaughlin (1999), these tiny and isolated systems are the objects most likely to have suffered considerable mass loss from the first round of supernovae, leaving behind whatever stars and clusters had managed to form before then. Using this picture, Dekel & Silk (1986) argue that the expected gas vs. stellar mass

Table 1.17. Cooling-flow galaxies in Abell clusters

Abell Cluster	BCG	Young GCs?	dM/dt
Virgo	M87	N	15
A426	N1275	Y	200
A496		N	38
A1060	N3311	Y?	10
A1795		Y	200
A2029	IC1011	N	260
A2052	U9799	(N)	55
A2107	U9958	(N)	8
A2199	N6166	(N)	70
A2597		N	135
MKW4	N4073	(N)	10

scalings for dwarfs should go as $(M_{gas}/M_\star) \sim L^{-0.4}$. But if the efficiency of cluster formation ϵ is constant (as was discussed above), then we should expect $S_N \sim L^{-0.4}$ for dwarf ellipticals. In deducing the role of the gas, note that there is one important difference between the dE's and the BCGs discussed above: in the dwarfs, the ejected fraction of the initial supply M_{gas} does not stay around in their halos. Only the stellar contribution $M_\star$ remains in their small dark-matter potential wells.

This scaling model ($S_N \sim L^{-0.4}$, starting at $M_T^V \simeq -18.4$ or about $2 \times 10^9 L_\odot$), is shown in Fig. 1.56. It does indeed come close to matching the observed trend for *nucleated* dE's but not the non-nucleated ones. At low luminosity ($M_V^T \sim -12$), we would expect from this scaling model that dE's should have S_N values approaching 20, much like what is seen in (e.g.) the Local Group dwarfs Fornax and Sagittarius, or some of the small nucleated dwarfs in Virgo. On the other hand, the non-nucleated dE's fall closer to the scaling model curve (1.60) which is simply the low-luminosity extension of the giant ellipticals. Intermediate cases are also present. Why the huge range between the two types?

The two brands of dwarf E's have other distinctive characteristics. The dE,N types have central nuclei which, in their spatial sizes and colors, resemble giant globular clusters (Durrell et al. 1996a; Miller et al. 1998) and indeed, some of the the smaller nuclei may simply be single globular clusters drawn in to the center of the potential well of the dwarf by dynamical friction (the cluster NGC 6715 at the center of the Sagittarius dE may be the nearest such example). The most luminous nuclei, however, far exceed even the brightest known globular clusters; these may represent true nuclei formed by strongly dissipative gaseous infall at a moderately early stage (e.g., Caldwell & Bothun 1987; Durrell et al. 1996a). It is also well known that the dE,N and dE types exhibit different spatial distributions in the Virgo and Fornax clusters (Ferguson & Sandage 1989): the dE,N types follow a more centrally concentrated distribution resembling the giant E's, while the dE types occupy a more extended distribution resembling the spirals and irregulars. Differences in shape have also been noted; the dE (non-nucleated) types have more elongated isophotes on average (Ryden & Terndrup 1994; Binggeli & Popescu 1995).

A plausible synthesis of this evidence (see Durrell et al. 1996a; Miller et al. 1998) is that the dE,N types represent "genuine" small ellipticals in the sense that they formed in a single early burst. Many of them clearly formed near the giant BCGs, and thus may have been in denser, pressure-confined surroundings which allowed some to keep enough of their gas to build a nucleus later (e.g., Babul & Rees 1992). The dE types, by contrast, may represent a mixture of gas-stripped irregulars, some genuine ellipticals, or even quiescent irregulars that have simply age-faded. For many non-nucleated dE's, the scaling model $S_N \sim L^{-0.4}$ involving early mass loss may not apply, and if there was much less mass loss then something closer to $S_N \sim const$ should be more relevant.

If this interpretation of the dwarfs has merit, then once again we must assume that the main era of globular cluster formation went on at a very early stage of the overall starburst, before the supernova winds drove out the rest of the gas. More direct evidence in favor of this view, such as from contemporary starburst dwarfs (see below), would add an important consistency test to this argument.

For the complete range of elliptical galaxies, the pattern of specific frequency with luminosity is shown in Fig. 1.56, with the McLaughlin model interpretation. It is encouraging that a plausible basis for interpreting the high$-S_N$ systems at both the top and bottom ends of the graph now exists, and that we have at least a partial answer to the classic "S_N problem". Nevertheless, individual anomalies remain at all levels, with cluster numbers that are too "high" or "low" for the mean line. Will we have to conclude from the high$-S_N$, non-BCG cases that genuinely high-efficiency cluster formation can indeed occur? Are all the low$-S_N$ cases just instances of simple gas-poor mergers of spirals, which had few clusters to begin with? There is much still to be done to understand these cases, as well as to fill in the complete story of early star formation in the central giant ellipticals.

1.9 FORMATION: MERGERS, ACCRETIONS, AND STARBURSTS

The way to get good ideas is to get lots of ideas and throw the bad ones away.

Linus Pauling

When we begin reviewing the issues relating galaxy formation to cluster formation, we are plunging into much more uncharted territory than in the previous sections. The flavor of the discussion must now shift to material that is less quantitative, and less certain. It is fair to say that in the past decade especially, we have isolated the *processes* that need to be understood in more detail (gas dynamics within the clumpy structure of protogalaxies, and later processes such as galaxy mergers, satellite accretions, galactic winds, tidal stripping, and dynamical evolution). However, we are not always able to say with confidence *which* of these mechanisms should dominate the formation of the GCS in any one galaxy. In this exploratory spirit, let us move ahead to survey the landscape of ideas as they stand at present.

The formation scenarios discussed in Section 8 can be classified as *in situ* models: that is, the galaxy is assumed to be formed predominantly out of an initial gas supply that is "on site" from the beginning. In such models,later modifications to the population from outside influences are regarded as unimportant. From the limited evidence now available, the *in situ* approach may well be a plausible one for many ellipticals, but it cannot be the whole story. We see galaxies in today's universe undergoing major *mergers*; *starbursts* from large amounts of embedded gas; and *accretions* of smaller infalling or satellite galaxies. Can these processes have major effects on the globular cluster populations within large spirals and ellipticals?

1.9.1 Mergers and the Specific Frequency Problem

Considerable evidence now exists that merging of smaller already-formed galaxies occurs at all directly visible redshifts. In a high fraction of these cases, the outcome of repeated mergers is expected to be an elliptical galaxy, and a traditional question is to ask whether all ellipticals might have formed this way. Considerable enthusiasm for the idea can be found in the literature over the past two decades and more. However, a primary nagging problem has to do with the specific frequencies of the relevant galaxies. Disk galaxies are, in this view, postulated to be the progenitors from which larger E galaxies are built. But disks or spirals consistently have specific frequencies in a rather narrow range $S_N \lesssim 2$, while (as discussed in Section 6 above) specific frequencies for ellipticals occupy a much larger range up to several times higher. This "S_N problem" is not the same one discussed in the previous section (which applied to the BCGs vs. normal ellipticals); instead, it addresses the offset in S_N between two very different types of galaxies. Trying to circumvent this problem – that is, making a high$-S_N$ galaxy by combining low$-S_N$ ones – has generated an interesting and vivid literature.

A brief review of the key papers in historical sequence will give the flavor of the debate. The seminal paper of Toomre (1977) first showed convincingly from numerical simulations that direct mergers of disk galaxies could form large ellipticals. Toomre speculated that a large fraction of present-day gE's originated this way. Not long after that, the first surveys of GCSs in Virgo and in smaller galaxy groups came available (Hanes 1977; Harris & van den Bergh 1981). Using this material, Harris (1981) suggested that the *low*$-S_N$ ellipticals, which are found characteristically in small groups and the field, might reasonably be argued to be the products of spiral mergers. Harris' paper concludes with the statement "The merger process cannot *increase* the specific frequency, unless vast numbers of extra clusters were somehow stimulated to form *during* a major collision early in its history, when substantial amounts of gas were still present". (It should be noted, however, that this last comment was a throwaway remark which the author did not really take as a serious possibility at that time!) Subsequently, van den Bergh (1982 and several later papers) repeatedly emphasized the difficulty of using disk-galaxy mergers to form "normal" (that is, Virgo-like) ellipticals in the range $S_N \sim 5$.

For descriptive purposes, let us call a *passive* merger one in which the stellar populations in the two galaxies are simply added together with no new star formation. In a passive merger, one would expect the specific frequency of the product to be the simple average of the two progenitors. Even after considerable age-fading of the Population I disk light, such a combination of low$-S_N$ disk galaxies would never yield a sufficiently high S_N to match the Virgo-like ellipticals.

The debate gained momentum when Schweizer (1987) emphasized that mergers of spirals could be *active* rather than passive: that is, the progenitors could contain considerable gas, and globular clusters could form *during the merger*, thus changing the specific frequency of the merger product. In Schweizer's words, "What better environment is there to produce massive clusters than the highly crunched gas in [merging] systems? ... I would predict that remnants of merged spirals must have more globular clusters per unit luminosity than the spirals had originally". Many subsequent authors took this statement as a signal that the "specific frequency problem" had therefore been solved. But it had not.

In two influential papers, Ashman & Zepf (1992) and Zepf & Ashman (1993) published a more quantitative model for the merger of gas-rich galaxies and active cluster formation during the merger, and used it to predict other characteristics (metallicities, spatial distributions) for the resulting GCS. Their formalism emphasized single merger events of two roughly equal spirals, though it could be extended to multiple events. A considerable stimulus for this model was the growing awareness that the metallicity distributions of the globular clusters in many giant ellipticals had a bimodal form, suggesting (in the Ashman-Zepf view) that the metal-richer population formed during the merger. In their argument, during the collision the gas from both galaxies would dissipate, funnel in toward the center of the new proto-E galaxy, and form new stars and clusters there. Many Gyr after the merger had finished, we would then see an elliptical with a bimodal MDF and with the metal-richer component more centrally concentrated.[11]

The Ashman/Zepf model also predicts that the metallicity gradient should be steeper for the GCS than for the halo light (the "younger" merger-produced MRC clusters would have formed preferentially in the core regions and with greater efficiency). This overall picture became additionally attractive with the discovery of "young globular clusters" (compact, cluster-sized star forming regions with masses extending up past the $\gtrsim 10^5 M_{\odot}$ range) in gas-rich interacting galaxies such as NGC 1275 (Holtzman et al. 1992) and NGC 3597 (Lutz 1991). Many similar cases are now known (discussed below), in which considerable gas seems to have collected by merger or accretion events.

The community response to this merger scenario was initially enthusiastic and somewhat uncritical. However, counterarguments were also raised. Even if globular clusters form during mergers, a higher specific frequency is *not* necessarily the result: field-star formation goes on at the same time as cluster formation, and the final S_N could be either higher *or* lower depending on the efficiency of cluster formation (Harris 1995).

It may seem attractive to assume that the highly shocked and compressed gas generated during disk mergers would be a good place for high-efficiency cluster formation. But it is not clear that these shocks would be any more extreme than in the range of collisions that took place in the protogalactic era, when much more gas was present and the random motions were comparably high. There seems no need to automatically assume that the cluster formation efficiency in *present-day mergers* would be higher than in the protogalactic era. There is, in addition, a problem of sheer numbers: in a normal Virgo-like elliptical there are thousands of MRC clusters, and the quantitative demands on the merger to create all of these are extreme (see Harris 1995 and the discussion below).

[11] Since Ashman & Zepf's original discussions, many authors have conventionally taken observations of bimodal MDFs as "supporting" the merger model. The correct statement is that bimodality is "consistent" with Ashman/Zepf. In general, observational evidence can be said to be *consistent* with a particular model if it falls within the expected results of that model. However, the same evidence might also be consistent with other models. To say that evidence *supports* a model is a much stronger statement: it requires the data to be consistent with that model *but inconsistent with other models*; that is, competing models are ruled out. We are fortunate indeed if our observations turn out to be strong enough to agree with only one model and to rule out competing ones! In this case, a bimodal MDF can equally well result from any *in situ* formation picture which involves at least two distinct epochs of star formation, with gaseous dissipation and infall occurring in between.

1.9.2 Observations of Merger Remnants

Clearly, what has been needed most of all to understand the characteristics of globular cluster formation during mergers is a series of new observations of merged galaxies. Whitmore, Schweizer, Zepf, and their colleagues have used the HST to carry out an important series of imaging studies of star clusters in galaxies that are clearly merger products, in an identifiable age sequence. Published cases include NGC 4038/39 where the young star formation is $\sim 10^8$ y old (Whitmore & Schweizer 1995); NGC 3256 at $\gtrsim 100$ Myr (Zepf et al. 1999); NGC 3921 (Schweizer 1996, Schweizer et al. 1996) and NGC 7252 (Miller et al. 1997), both at an age $\sim$ 700 Myr; and NGC 1700 and 3610 (Whitmore et al. 1997), at ages of 3 to 4 Gyr. Other studies in this series are in progress. These galaxies are excellent testbeds for making quantitative measurements of young vs. old cluster subpopulations, asking where they are in the merging material, and deriving their specific frequencies and mass distributions.

An example of the data from the NGC 7252 study is shown in Fig. 1.57. Two groups of clusters are clearly visible, with the brighter, bluer and spatially more centrally concentrated population identified as the objects formed in the merger. In this case, the color difference between the bright, blue clusters and the fainter, redder ones is interpreted as primarily due to age (750 Myr vs. $\gtrsim 10$ Gyr) rather than metallicity. One can, however, make reasonable estimates of how the younger population would evolve in luminosity and color as the galaxy ages ("age fading"). In the Whitmore/Schweizer papers the stellar population models of Bruzual & Charlot (1993) are used, in which the clusters are assumed to form in a single burst and simply evolve passively by normal stellar evolution after that. Whitmore et al. (1997) present an interesting numerical simulation showing how a single-burst population of clusters would evolve progressively in a color-magnitude diagram such as Fig. 1.57. Intriguingly, for roughly solar abundance and ages near $\sim 1-3$ Gyr, the integrated colors of the clusters are near $(V-I) \simeq 1$, much like conventional old-halo globular clusters that are metal-poor ([Fe/H] $\simeq -1.7$). Thus for intermediate-age mergers, separating out the two types of cluster populations becomes extremely difficult.

For younger mergers, the mean age of the burst can be plausibly estimated by the color of the blue clusters and by the velocities and geometry of the progenitors, if they are still separate (Whitmore & Schweizer 1995; Fritze-von Alvensleben 1998). From the age-fading models, we can then estimate the LDF of the young clusters (number per unit luminosity) as it would look at a normal old-halo age. In Fig. 1.58, the age-faded LDFs for NGC 4038/39 and NGC 7252 are compared directly with that of M87. For the upper range $L \gtrsim 0.5 \times 10^5 L_\odot$, all three galaxies match extremely well, confirming our earlier suggestions that the mass distribution function is not strongly affected by dynamical evolution for $M \gtrsim 10^5 M_\odot$. However, at lower masses, the M87 curve diverges strongly from the other two, falling well below them.

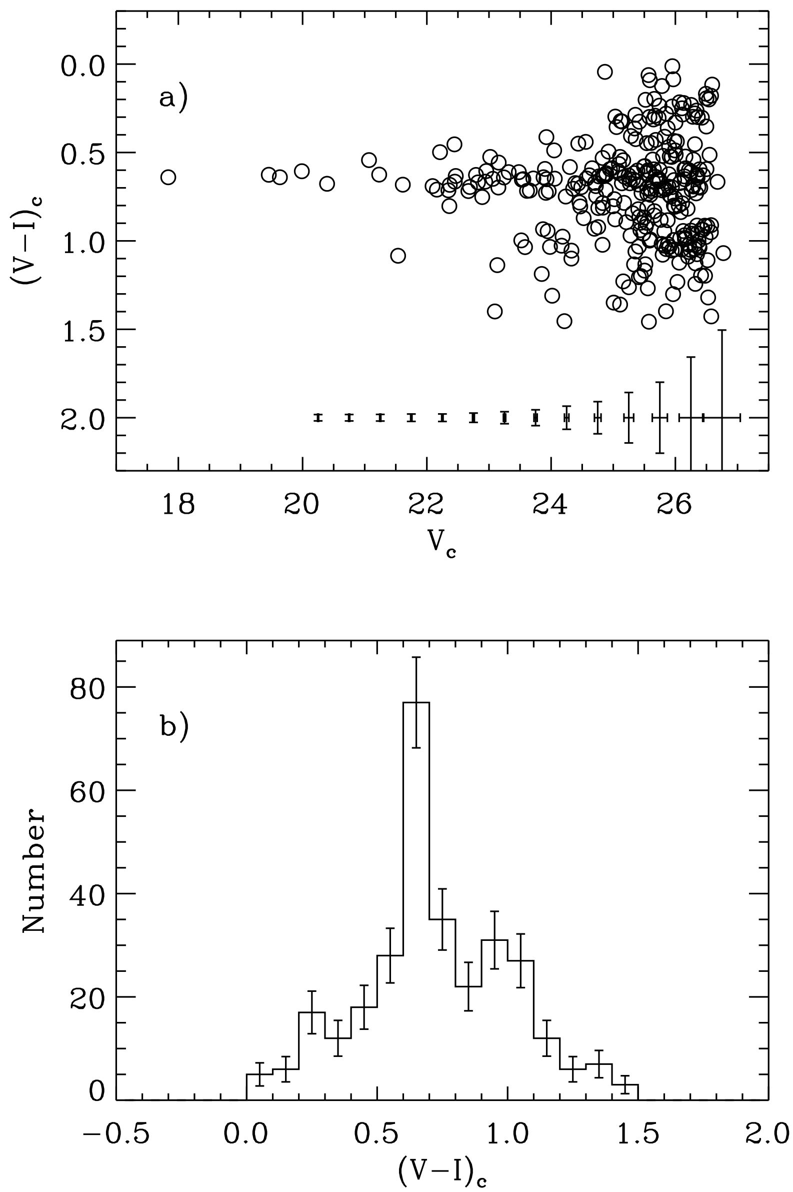

Fig. 1.57. (a) Upper panel: integrated colors $(V - I)$ and apparent magnitudes V for the star clusters in the merger product NGC 7252 (Miller et al. 1997). The brighter, more numerous clusters centered at a mean color $(V - I) \simeq 0.65$ are presumed to be ones formed in the merger, while the redder, fainter population centered at $(V - I) \simeq 1.0$ is likely to be the old-halo cluster population from the two original galaxies. (b) Lower panel: Histogram of cluster colors. The two modes (young vs. old) are clearly visible. Figure courtesy Dr. B. Miller

Is this a signal that $\sim 10^{10}$ years of dynamical evolution have "carved away" this low-mass end of the LDF? Or have the numbers of low-mass clusters in the merger remnants been overestimated by observational selection effects? In Section 8 above, it was noted that collisional growth model does predict an LDF shape continuing upward to low masses much like the observations, but this match is, perhaps, based on too simple a set of model assumptions.

In general, however, the LDF shapes in these obvious merger remnant galaxies closely match what is expected from the collisional-growth theory, $dn/dL \sim L^{-\alpha}$ with $\alpha = 1.8 \pm 0.2$ (assuming constant M/L). In NGC 4038/39 (the Antennae system), the slope is $\alpha = 1.78 \pm 0.05$; in NGC 3256, we have $\alpha = 1.8 \pm 0.1$; in NGC 3921, $\alpha = 2.12 \pm 0.22$; and in NGC 7252, $\alpha = 1.90 \pm 0.04$.

As for the sites of cluster formation, Whitmore & Schweizer (1995) note for the Antennae (the youngest merger) that "many of the clusters form tight groups, with a single giant HII region containing typically a dozen clusters". This clumpy distribution is exactly what we would expect from the SGMC formation picture, where large amounts of gas must be collected together to form local reservoirs from which cluster formation can proceed. Another intriguing feature of the Antennae system is that star formation is occurring at a high rate *well before the progenitor galaxies have completely merged.* The nuclei of the original disk galaxies are still clearly visible, and the disk gas has obviously not waited to "funnel" down in to the merged nucleus before starting star formation in earnest. Shocked, clumpy gas appears all over the merger region and even out along the tidal tails.

1.9.3 A Toy Model for Mergers and Specific Frequencies

Many of the issues surrounding the effect of mergers on specific frequencies can be clarified by building a simple quantitative model. Let us assume that two initial galaxies with luminosities (L_1, L_2) and specific frequencies (S_{N1}, S_{N2}) will merge to form an elliptical. (NB: these values are assumed to be "age-faded" ones, i.e. where all the light of the galaxy is reduced to the level it would have at an old-halo age $\sim 10 - 15$ Gyr.) Assume further that the galaxies bring in a total amount of gas M_g which is turned into new stars and clusters. There may be *additional* gas which is left unused; here we simply assume M_g is the amount actually turned into stars.

How many globular clusters do we get? During the merger we will form N_3 new clusters with a total mass M_{cl} at an efficiency (using our previous notation)

$$\epsilon = \frac{M_{cl}}{M_g} = 3 \times 10^5 M_\odot \frac{N_3}{M_g} \tag{1.61}$$

where the mean cluster mass is $3 \times 10^5 M_\odot$, assumed the same as the Milky Way. We will *also* form new stars, with a total (age-faded) luminosity

$$L_3 = \frac{M_g}{(M/L)_V} \tag{1.62}$$

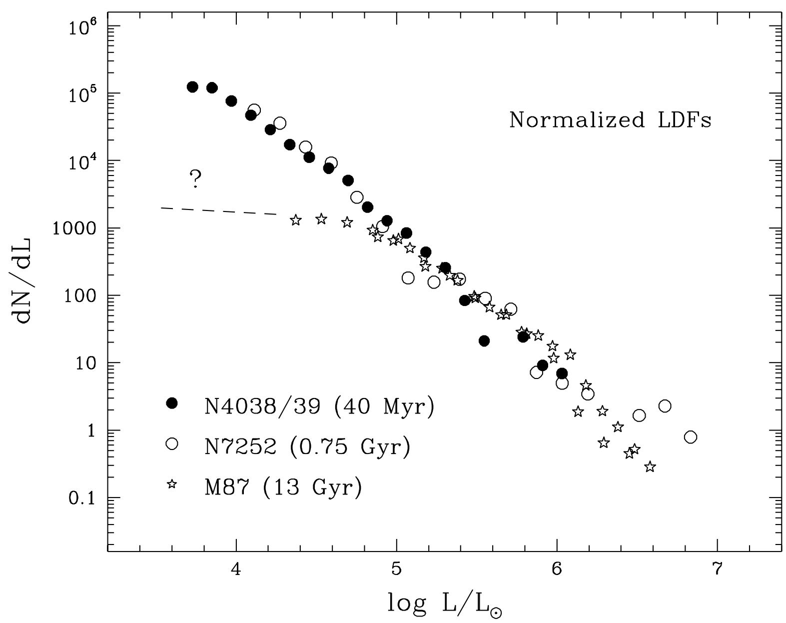

Fig. 1.58. Luminosity distribution functions (LDFs) for three galaxies: NGC 4038/39 (the Antennae merger system, at an age 40 Myr), NGC 7252 (a 750 Myr merger), and M87. The two merger remnants have been age-faded to an equivalent age of 13 Gyr with the Bruzual-Charlot models (see text) and then superposed on the M87 data for comparison

where $(M/L)_V \simeq 8$ for old (Population II) stellar populations. Using $S_N = 8.55 \times 10^7 (N_{cl}/L)$, we can quickly show that our "baseline" specific frequency $S_N^0 \simeq 3.5$ corresponds to an efficiency $\epsilon_0 = 0.0015$. (This is about a factor of two less than for the Virgo and Fornax giant ellipticals analyzed by McLaughlin 1999, which have $S_N \sim 5$).

Now we add up the old and new clusters to get the final specific frequency in the product elliptical:

$$S_N(\text{final}) \; = \; 8.55 \times 10^7 \left(\frac{N_1 + N_2 + N_3}{L_1 + L_2 + L_3} \right) \tag{1.63}$$

The L_3 term is forgotten by many writers.

Whether or not S_N ends up higher or lower than (S_{N1}, S_{N2}) clearly depends on both N_3 and L_3. Rewriting the result in terms of the formation efficiency ϵ and input gas mass M_g, we obtain

$$S_N \; = \; \left(\frac{S_{N1} L_1 + S_{N2} L_2 + 0.44 M_g (\epsilon/\epsilon_0)}{L_1 + L_2 + M_g/8} \right) \tag{1.64}$$

This is the general case for any two-galaxy merger. We can obtain a better idea of the effects if we look at specific cases. A particularly important one is for *equal progenitors*, that is, $L_1 = L_2$. If the merger is "passive" (gas-free; $M_g/M_{1,2} \ll 1$), then clearly

$$S_N = \frac{1}{2}(S_{N1} + S_{N2}) \ . \tag{1.65}$$

This latter situation is the limiting case discussed by Harris (1981). It was used to justify the suggestion that low$-S_N$ ellipticals might be the end products of spiral mergers, especially in small groups where the galaxy-galaxy collision speeds are low, enhancing the probability of complete mergers.

The range of possibilities for "active" (gas-rich), equal-mass mergers is illustrated in Fig. 1.59. This graph shows an Antennae-like merger; both the incoming galaxies are adopted to have $M_1 \simeq M_2 = 1.7 \times 10^{11} M_\odot$, equivalent to $M_V \simeq -21$. We assume them both to have age-faded specific frequencies $S_N \simeq 3$ (already an optimistically high value for spirals). For an input gas mass of $10^{10} M_\odot$ (that is, about 3% of the total mass of the progenitors), the model show only modest changes in the outcome S_N are possible even if the cluster formation during the merger is enormously efficient: there is simply not enough gas in this case to build a significant number of new clusters. Nevertheless, it is a plausible source for a low$-S_N$ elliptical.

For input gas masses $M_g \gtrsim 5 \times 10^{10} M_\odot$ (15% or more of the total galaxy mass, which corresponds to quite a gas-rich encounter), larger changes in S_N are possible *but only if the cluster formation efficiency is far above normal.* The essential prediction of this toy model is, therefore, that the expected result of a major merger will be an elliptical with $S_N \sim 3$, unless there is a *huge amount of input gas* and *a very high cluster formation efficiency*; both conditions must hold. Only then can we expect to build a Virgo-like (or, even more extreme, a BCG-like) elliptical this way.

For comparison, the Antennae merger has $\sim 2 \times 10^9 M_\odot$ of molecular gas (Stanford et al. 1990), while the extremely energetic starburst system Arp 220 has $\sim 10^{10} M_\odot$ of H_2 (Scoville 1998). These are insufficient amounts of raw material to generate major changes in the specific frequencies. The most extreme case that may have been observed to date is in the giant starburst and cD galaxy NGC 1275. There, large amounts of gas have collected in its central regions, and many hundreds of young clusters or cluster-like objects have formed (Holtzman et al. 1992; Carlson et al. 1998). Carlson et al. estimate that, of the ~ 1180 young blue objects detected in their study (most of them of course at low luminosity), perhaps half would survive after 13 Gyr, assuming that they are all indeed globular clusters. Even this large new population, though, is not capable of changing the global specific frequency of this galaxy noticeably from its current level $S_N \sim 10$. Furthermore, spectra of five of the brighter young cluster candidates (Brodie et al. 1998) call into question their identification as globular clusters. Their integrated spectral properties are unlike those of young Magellanic or Galactic clusters, and can

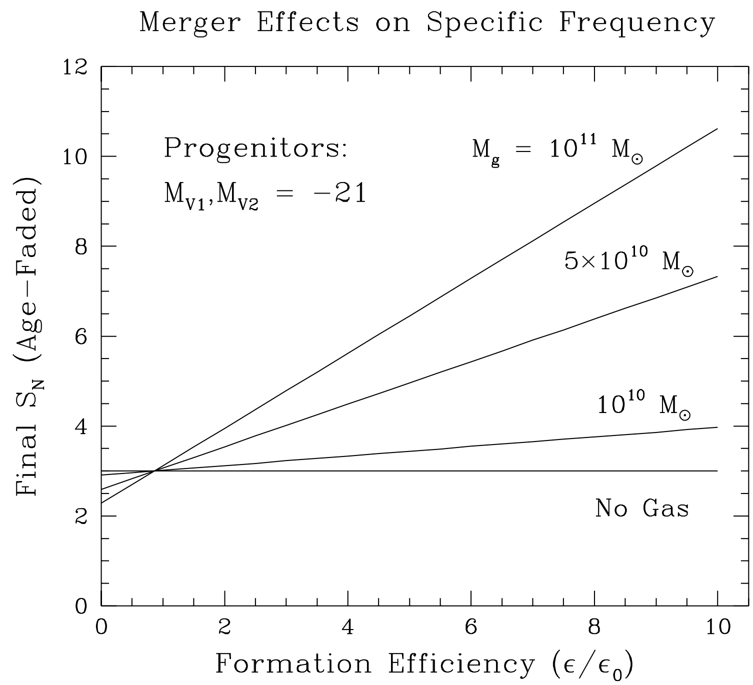

Fig. 1.59. Specific frequency S_N for an elliptical galaxy formed from the merger of two equal spirals. Here S_N is plotted versus the efficiency of cluster formation ϵ relative to the "normal" efficiency $\epsilon_0 = 0.0015$ (see text). The four curves are labelled with the amount of gas M_g converted into stars during the merger

be interpreted as clusters with initial mass functions weighted strongly to the high-mass end. If so, a high fraction of them may self-disrupt or fade quickly away to extinction. This material emphasizes once again that we need to understand the first ~ 1 Gyr of evolution of a GCS before we can properly calculate the effects on the specific frequency and LDF.

The expected range of S_N can be compared with the actual merger products studied by Whitmore, Schweizer and their colleagues referred to above. In each case they have made empirical estimates of the global S_N in the end-product elliptical, and in each case it is at a level $S_N \lesssim 3$ consistent with the view that cluster formation efficiency during the merger is, in fact, not much different from the normal ϵ_0 level.

One of these calculations will illustrate the technique: for the Antennae, Whitmore & Schweizer (1995) find a total of ~ 700 young, blue objects (assumed to be mostly clusters); however, they identify only 22 to be brighter than the classic GCLF turnover point at $\sim 10^5 M_\odot$ and these are the important ones for calculating the specific frequency long after the merger is over, since the small or diffuse clusters will have largely been destroyed. Adding these to an estimated $\sim 50 - 70$ old-halo clusters already present in the progenitor galaxies then gives a total population of $\lesssim 100$ old clusters and a total (age-faded) E galaxy luminosity $M_V^T \simeq -20.7$ long after the merger is complete. The resulting specific frequency is $S_N \simeq 0.5$, which places it at the bottom end of the scale for observed E galaxies. The more recent analysis of the same material by Fritze-von Alvensleben (1998, 1999), who uses individual age estimates for each object to estimate their masses, suggests instead that the number of young clusters more massive than the mass distribution is already Gaussian in number per unit log mass, and that there may be as many as ~ 150 above the turnover. Adopting this higher total, however, only raises the final estimate to $S_N \simeq 1.2$. In short, the Antennae merger is producing a cluster-poor elliptical.

Similar calculations for the other galaxies in the series listed above (NGC 1700, 3610, 3921, 5018, 7252) are perhaps a bit more reliable since they are older remnants, and differences in internal reddening and age are less important. For these, the age-faded specific frequencies predicted empirically (see the references cited above) are all in the range $1.4 \lesssim S_N \lesssim 3.5$. The model displayed in Fig. 1.59 with normal production efficiencies appears to be an entirely tolerable match to these observations.

The evidence from specific frequencies is clear that E galaxies can be built, and are being built today, by mergers of pre-existing disk systems. But the type of elliptical being produced in such mergers resembles the low$-S_N$ ones in small groups and in the field.

1.9.4 Other Aspects of the Merger Approach

The ellipticals in rich clusters provide a much stronger challenge to the simple merger scheme. The first and perhaps biggest barrier is connected with the sheer numbers of clusters in these gE's, and can be illustrated as follows (Harris 1995). Let us take NGC 4472 in Virgo as a testbed "normal" object ($S_N \simeq 5$). Its clusters display a bimodal MDF with about 3660 in the metal-poor population and 2400 in the metal-richer population. Now, *if* the entire galaxy formed by mergers, then by hypothesis all the MPC clusters must have come from the pre-existing disk systems. This would require amalgamating about 30 galaxies the size of the Milky Way, or an appropriately larger number of dwarfs (see below). Similarly, suppose we assume optimistically that all the MRC clusters formed during the mergers at rather high efficiency ($\epsilon \simeq 0.003$); then the amount of input gas needed to do this would be at least $2.4 \times 10^{11} M_{\odot}$, or $\sim 10^{10} M_{\odot}$ per merger. These are *very* gas-rich mergers. The amounts may be even larger if we account for wastage, i.e. gas lost to the system. In addition, over many different mergers it is not clear that a cleanly bimodal MDF could be preserved.

Alternately, one could assume that the galaxy was built by just one or two much larger mergers with almost all the gas ($\sim 3 \times 10^{11} M_{\odot}$) coming in at once. The progenitor galaxies would, in fact, then have to be mostly gaseous, unlike any merger happening today. The only epoch at which such large supplies of gas were routinely available was the protogalactic one. It is then not clear how the merger scheme would differ in any essential way from the regular *in situ* (Searle/Zinn-like) model.

Specific frequencies are not the only outcome of a merger that is amenable to observational test. Detailed evaluations of other measurable features, particularly the metallicity distributions, have been given recently especially by Geisler et al. (1996), Forbes et al. (1997), and Kissler-Patig et al. (1998). Noteworthy problems that arise from these analyses are as follows:

- In the merger model, the highest$-S_N$ ellipticals should have the great majority of their clusters in the metal-rich component, since by hypothesis these are created during the merger at high efficiency. A few of them do (NGC 3311 in Hydra I and IC 4051 in Coma; see Secker et al. 1995, Woodworth & Harris 1999), but many certainly do not (notably M87 and NGC 1399, where the metal-poor clusters are in a slight majority).
- If a sequence of mergers is required to build up a giant elliptical, then the gas – which is enriched further at each stage – should produce a multimodal or broad MDF, rather than the distinct bimodal (or even narrow, unimodal) MDFs that are observed.
- In most giant ellipticals, the metal-poor component is itself at *higher* mean metallicity (typically [Fe/H] $\simeq -1.2$; see Forbes et al. 1997) than the MPCs in spirals and dwarfs (at [Fe/H] $\simeq -1.6$), suggesting that they did not originate in these smaller systems after all.
- Radial metallicity gradients in the GCSs are usually *shallower* in ellipticals with *lower* S_N, the reverse of what is expected from the merger model.

A still newer line of attack, complementary to the specific frequencies and MDFs, is in the kinematics of the cluster systems. With the new 8- and 10-meter telescopes, it is now possible to accumulate large samples of accurate radial velocities for the globular clusters in the Virgo and Fornax ellipticals and thus to carry out kinematical analyses similar to the ones traditionally done for the Milky Way or M31. Data of this type for M87 (Kissler-Patig & Gebhardt 1998; Cohen & Ryzhov 1997) suggest, albeit from sketchy coverage of the halo at large radii, that the *outer* part of the halo ($R_{gc} \gtrsim 20$ kpc, mostly from the MPC clusters) shows a substantial net rotation of 200 km s^{-1} or more, directed along the isophotal major axis. The MRC clusters by themselves show a more modest net rotation $V_{rot} \simeq 100$ km s^{-1} at all radii. Was this the result of a single merger between two galaxies that were already giants? This would be a simple way to leave large amounts of angular momentum in the outer halo. However, such a conclusion does not seem to be forced on us; even the merger product of several large galaxies can yield large outer-halo rotation (e.g., Weil & Hernquist 1996). In the other Virgo supergiant, NGC 4472, the cluster velocities (Sharples et al. 1998) confirm that the MRC has a distinctly lower velocity dispersion and is thus a dynamically cooler subsystem. In addition, there are hints from their still-limited dataset that the MPC has a distinctly higher net rotation speed, in analogy with M87. The interpretation of these systems is growing in complexity, and they present fascinating differences when put in contrast with the Milky Way.

In summary, mergers undoubtedly play a role in the formation of large ellipticals. They may be the dominant channel for forming ones in small groups, out of the S and Irr galaxies that are found in large numbers in such environments. However, the bigger and higher$-S_N$ ellipticals in rich environments present several much more serious problems, which do not appear to be met by the Ashman/Zepf scenario in its initial form.

1.9.5 The Role of Accretions

The "merger" of a small galaxy with a much larger one is normally called an *accretion*. This is another type of event which must be fairly common for any large galaxy with significant numbers of satellites (again, the absorption of the Sagittarius dE by the Milky Way is the nearest and most well known example). The possibility of building up the observed GCS characteristics in giant ellipticals by accretions of many small satellites was pursued in some early numerical simulations by Muzzio (1986, 1988, and other papers; see Harris 1991 for an overview). It has been investigated anew by Côté et al. (1998) with the specific goal of reproducing the observed metallicity distribution functions. In brief, the assumption of their model is that an original central elliptical forms in a single major phase, giving rise to the MRC clusters. The correlation of the mean [Fe/H](MRC) with galaxy luminosity noted by Forbes et al. (1997) is laid down at this time (the bigger the galaxy, the more metal-rich the MRC component is).

Then, the "seed" galaxy – large, but not nearly at its present-day size – begins to accrete neighboring satellites. These are drawn randomly from a Schechter galaxy luminosity function, so the majority are dwarfs. Since all the accreted objects are smaller, their attendant globular clusters are more metal-poor, and over time, the entire MPC component builds up from the accreted material. In the sense used above, this process is assumed to be a *passive* one, with no new star formation. The Côté et al. model provides a valuable quantification of the results to be expected on the MDF from accretion. The attractive features are that it has the potential to explain the wide galaxy-to-galaxy differences in the MDF for the metal-poor component (Forbes et al.), while maintaining the similarity of the metal-rich clusters from one gE to another. In one sense, the accretion model is the reverse of the merger model: in the accretion scenario, the MRC clusters are the ones belonging to the "original" gE, while in the merger picture, the MRC clusters are formed actively during the buildup.

The Virgo giant NGC 4472 again is used as a template for specific comparisons. Some of these are shown in Fig. 1.60. Two obvious points to be drawn from the comparison (see also others shown in Côté et al.) are that the model has the flexibility to produce a wide range of bimodal-type MDFs; and that *large* numbers of dwarfs – many hundreds – need to be accreted to build up the metal-poor component sufficiently. The steeper the Schechter function slope, the larger the proportion of globular clusters accreted from small dwarfs and the more metal-poor the MPC appears.

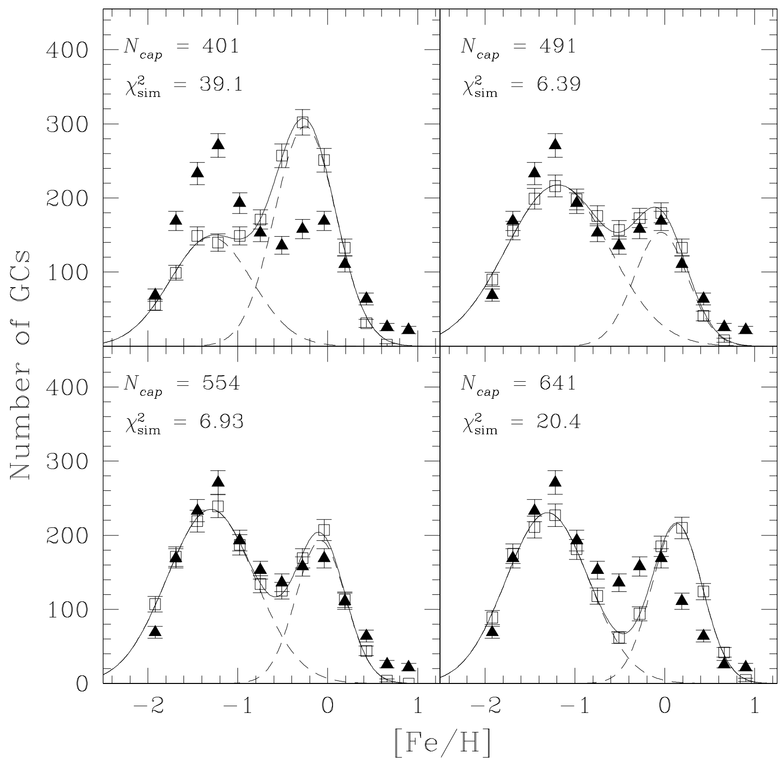

Fig. 1.60. Four synthetic metallicity distribution functions for globular clusters in a giant elliptical, taken from Côté et al. (1998). The metal-poor part of the bimodal MDF is produced from N_{cap} captured dwarf galaxies which were drawn randomly from a Schechter LF with slope −1.8. Open squares represent the final model galaxy; closed triangles show the observed MDF for NGC 4472 in Virgo; and the lines are the double-Gaussian combination best fitting the synthetic galaxy. Figure courtesy Dr. P. Côté

The passive-accretion model has the distinct advantage of using a process that must surely happen at some level in an ongoing fashion. It does, however, have its share of characteristic problems:

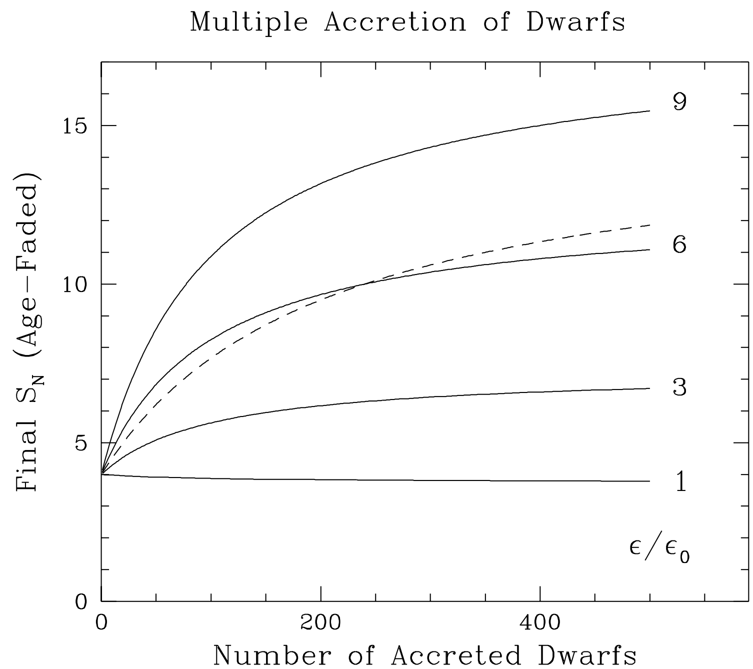

Fig. 1.61. Specific frequency S_N for an elliptical galaxy which grows by the accretion of many gas-rich dwarf galaxies. The initial E galaxy is assumed to have $M_V = -21.2$ ($M \sim 2 \times 10^{11} M_\odot$) and has specific frequency S_N(init) = 4. The accreted dwarfs each have $10^9 M_\odot$ of stellar mass and $10^9 M_\odot$ of gas mass, all of which is converted to stars and globular clusters during the accretion. The relative efficiency of globular cluster formation ϵ/ϵ_0 is denoted at right. The *dashed line* indicates the result of "passive" accretion of gas-free dwarfs (no new star formation) which all have $S_N = 15$

- If $\sim$ 500 small galaxies were needed to build up (say) NGC 4472, then many thousands of them would have been absorbed over the whole Virgo cluster to generate the many giant ellipticals there today. This would require, in turn, that *most* (perhaps 90%) of the original dwarf population is now gone. Is this plausible? Perhaps. However, the collision cross sections for dwarf ellipticals are small, and full-scale dynamical simulations may be needed to test the expected accretion rates in detail (it is interesting to note that the earlier semi-analytical studies by Muzzio predicted rather small exchange effects from the dwarfs).
- The specific frequencies of spiral galaxies (suitably age-faded) and many dE's are in the range $S_N \sim 1 - 3$, whereas the present-day Virgo and Fornax gE's have $S_N \simeq 5$. Reconciling these figures would require us to

assume that the initial metal-rich seed elliptical had $S_N \simeq 8$, which is in the BCG range – not a fatal objection, but one which does not necessarily favor the model. Côté et al. suggest that the gE may avoid diluting its cluster population this way if it accretes only the outer envelopes of the incoming galaxies by tidal stripping. Since the globular clusters in dE galaxies are found preferentially in the outskirts of their galaxies, the accreted material would be cluster-rich and S_N could stay nearly constant or even increase. But if this were the case, then the present-day gE's should be surrounded by hundreds (or thousands, in the case of M87) of stripped cores of former dE's. These remnants should be easily noticeable, but where are they?

- The most worrisome problem seems to me to be connected with the metallicity of the halo light. In this model, roughly *half* of the gE is accreted, low-[Fe/H] material. Thus, the mean color of the halo light should be roughly halfway between the red MRC and the bluer MPC. However, the measured color profiles of gE halos *match extremely well with the color of the MRC clusters* (see Geisler et al. 1996 and Fig. 1.47 for NGC 4472). How can such a galaxy accrete metal-poor clusters without accreting almost no field stars?

Another approach to the accretion scenario is taken by Kissler-Patig et al. (1999), who use the measured specific frequencies in the Fornax giant ellipticals to argue that the cD NGC 1399 might have accreted clusters and halo material predominantly from the neighboring *large* ellipticals in Fornax rather than dE's. This approach would, at least, circumvent the metallicity issue mentioned above. However, full-scale dynamical simulations will be needed to evaluate the plausibility of this level of stripping and halo redistribution among the large galaxies.

Passive accretion therefore leaves some fairly serious difficulties, just as did passive mergers. However, the conditions would change significantly if the accreted small galaxies were highly gaseous, so that (once again) new clusters could form in the process (see also Hilker 1998 for a similar view). This *active accretion* could then allow most of the field stars in the merged product to be at the necessary high metallicity, since most of them would have formed in the accretions. Clusters would also be added to both the MRC and the original MPCs that were present in the seed elliptical and the accreted dwarfs. However, just as for the merger picture, the entire formation model would once again begin to resemble a Searle/Zinn-like one whereby the galaxy builds up from many small gas clouds.

The "toy model" for mergers described in the previous section can be used to evaluate the effect of multiple accretions as well. In Fig. 1.61, the final S_N of the product gE is plotted as a function of the number of accreted dwarfs. The initial E galaxy is assumed to be a moderately luminous elliptical but not a giant ($M_V = -21.2$), while the accreted dwarfs have equal amounts of gas and stars ($10^9 M_\odot$ each; these are *very* gas-rich dwarfs). All the gas is assumed to be converted to stars during the accretions, with a cluster formation efficiency ϵ as defined previously. It is clear that, to attain final S_N values in the BCG range ($\gtrsim 10$), one needs to accrete many hundreds of dwarfs and convert their gas to clusters with abnormally high efficiency.

Alternately, one can assume that the accretions are "passive" (no new gas or star formation) and that the accreted dE's themselves have high $S_N \sim 15$. The result is shown as the dashed line in Fig. 1.61, and is roughly equivalent to the case for $\epsilon/\epsilon_0 = 6$.

1.9.6 Starburst Galaxies

To gain additional help in understanding what the early stage of galaxy formation from many small, dwarf-sized pieces may have looked like, we should take a closer look at the small galaxies in which large amounts of star formation are now taking place. Massive star clusters (young globular clusters) are seen to be forming in many individual galaxies which have large amounts of gas and are undergoing energetic star formation at the present time. Such galaxies are loosely called "starburst" systems, and their embedded young clusters have often been called "super star clusters" in the literature. Such clusters were known to exist more than 20 years ago (e.g., in NGC 1569 and 1705; see below), but were not connected until recently with globular clusters. This unfortunate communication gap persisted because the community of astronomers studying these young galaxies, and the community studying traditional globular clusters, had little contact with each other. The traditional and needlessly restrictive paradigm of globular clusters as exclusively old objects has taken a long time to fade away.

The small, dwarf-sized starburst galaxies are extremely interesting laboratories for our purposes. Many of these have high proportions of gas, and promise to give us our best direct view of what the protogalactic SGMCs (Searle/Zinn fragments) may have looked like. One of the nearest and best studied of these, NGC 5253 in the Centaurus group, is shown in Fig. 1.62 (Calzetti et al. 1997). If we could visualize many dozens of these dwarfs, sprinkled across a $\sim 50-$kpc region of space and all undergoing their first starbursts, we might gain an image of what the early Milky Way galaxy looked like.

Many starburst dwarfs are found to have handfuls of young star clusters (or, at the very least, associations) that are massive enough to qualify as genuine young globular clusters. A summary of several of them, drawn from the recent literature, is shown in Table 1.18. Here, columns (1) and (2) give the galaxy name and the number of young, massive star clusters in it; and column (3) the age of the starburst in Myr, usually estimated from such factors as the integrated colors of the clusters, the luminosities and colors of the OB field stars present, and the presence or absence of HII regions and Wolf-Rayet stars. Columns (4-6) give the total mass M_{cl} in all the massive young clusters, the estimated mass $M(H_2)$ in molecular hydrogen contained

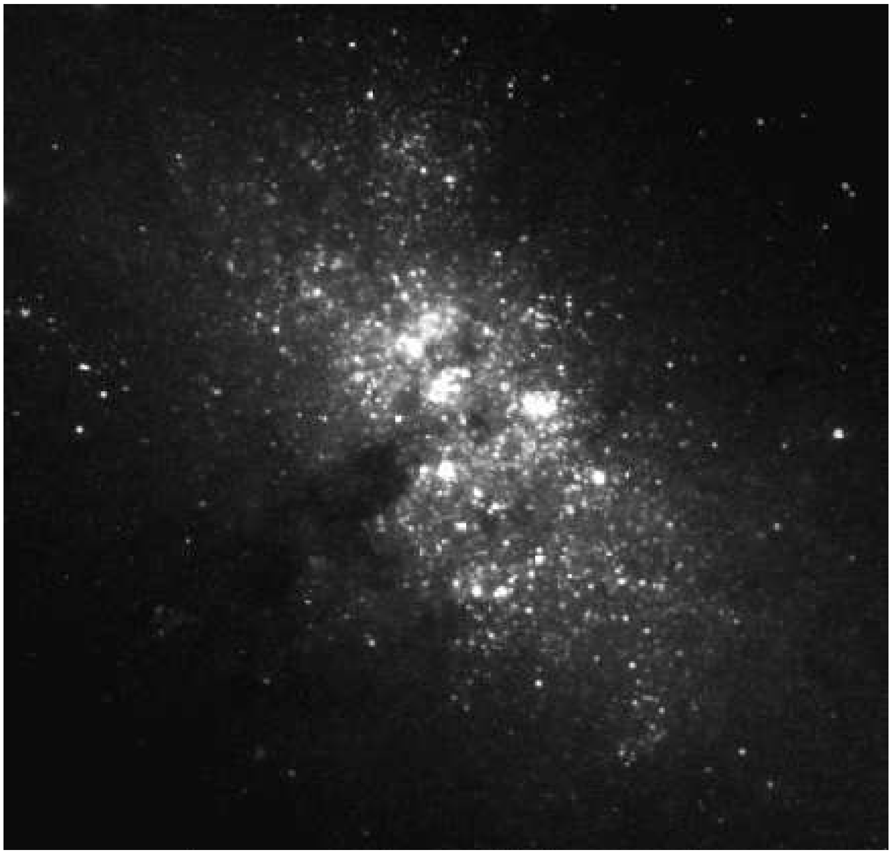

Fig. 1.62. *HST* image of the starburst dwarf galaxy NGC 5253. The field of view shown is about 0.9 kpc across, with the assumption that the galaxy is 4 Mpc distant. Young OB stars are seen all across the face of the dwarf, along with half a dozen brighter clumps in the central regions, which appear to be massive young star clusters. Image courtesy of Dr. D. Calzetti

within the dwarf, and the mass $M(HI)$ in neutral hydrogen. All masses are given in $M_{\odot}$. (NB: The last entry, NGC 4449, is a fairly large system resembling the LMC in many respects and it is not easily possible to assign a single age to the star formation epoch.) Many others are known in addition to the ones listed, though with less complete material for the embedded star clusters (see, e.g., Meurer et al. 1995; Mayya & Prabhu 1996).

There are several common themes to be drawn out of this comparison:

- Massive cluster formation takes place in these little starburst systems preferentially in the *central, densest* regions of the collected gas.
- Many of these dwarfs are rather isolated systems, and their recent starburst has not obviously been provoked by tidal encounters or other exter-

Table 1.18. Starburst dwarf galaxies with young clusters

Galaxy	N_{cl}	Age (Myr)	M_{cl}(tot)	$M(H_2)$	$M(HI)$	Source
NGC 1140	12	10	8×10^5		$10^8 - 10^9$	1,2
NGC 1569	7	5	6.4×10^5	5×10^7	1.4×10^8	3,4,5,6
NGC 1705	2	13	$4, \times 10^5$		1.2×10^8	5,7
NGC 4214	few	5	$\gtrsim 10^5$	1.0×10^8	1.1×10^9	8,9
NGC 5253	6	$\lesssim 10$	$\sim 10^6$	$\lesssim 2 \times 10^8$		10,11,12,13
He 2-10	19	$\simeq 5$	$10^6 - 10^7$	1.6×10^8	2×10^8	14,15,16
UGC 7636	18	10; 100	5×10^5		7×10^7	17
NGC 4449	dozens			8.7×10^8	5×10^9	18,19,20

Sources: (1) Hunter et al. 1994a (2) Hunter et al. 1994b (3) De Marchi et al. 1997 (4) Gonzalez Delgado et al. 1997 (5) O'Connell et al. 1994 (6) Waller 1991 (7) Meurer et al. 1992 (8) Leitherer et al. 1996 (9) Kobulnicky & Skillman 1996 (10) Calzetti et al. 1997 (11) Turner et al. 1997 (12) Beck et al. 1996 (13) Gorjian 1996 (14) Conti & Vacca 1994 (15) Kobulnicky et al. 1995 (16) Matthews et al. 1995 (17) Lee et al. 1997 (18) Bothun 1986 (19) Tacconi & Young 1985 (20) Israel 1997

nal stimuli. No outside "trigger" is apparently required to set off vigorous star formation.
- Many of the starbursts listed above are extremely young ($\lesssim 10$ Myr). Give or take a few Myr, the data indicate that the clusters form contemporary with the field stars (the "distributed" blue light), or at least during the leading edge of the burst.
- Large amounts of gas are present, and the burst does not appear to consume the entire supply. The residual amount of HI and H_2 gas not yet converted into stars is always $\gtrsim 100$ times more than the total mass contained in the young clusters.

All of these factors are consistent with the characteristics of the formation model outlined in Section 8: the available reservoir of gas must be at least $10^8 M_\odot$ to form globular-sized clusters: a handful of clusters form within one SGMC; their total masses use up only $\lesssim 1\%$ of the total gas supply; and they form at an early stage of the burst.

Massive star clusters are also seen forming in much larger galaxies – again, always as part of an energetic starburst event. In these cases, the source of the burst may be an accretion of a gas-rich satellite (thus not a "merger" in the restricted sense used in the previous section), a tidal shock from a close encounter, or the collection of gas into a central ring or bar, among other mechanisms. The most well known of these cases are probably NGC 1275 (Holtzman et al. 1992; Carlson et al. 1998) and NGC 3597 (Lutz 1991; Holtzman et al. 1996), thanks to the recent high resolution imaging of the HST cameras which has revealed many details of the nuclear star formation activity in these distant systems. A summary of parameters – numbers and masses of young clusters and total gas mass – for several large starburst galaxies is given in Table 1.19. As above, masses are given in $M_\odot$ units.

Table 1.19. Large starburst galaxies with young clusters

Galaxy	N_{cl}	Age (Myr)	M_{cl}(tot)	$M(H_2)$	$M(HI)$	Source
NGC 1275	~ 1180	$\gtrsim 100$	$> 10^8$	1.6×10^{10}	$\gtrsim 5 \times 10^9$	1,2,3,4,5
M82	> 100	10:	$\gtrsim 10^6$	2×10^8	2×10^8	6,7,8
NGC 253	4	$10 - 50$	2×10^6	2×10^8	4×10^8	9,10,11
NGC 5128	dozens	50:	10^5:	3×10^8	3×10^8	12,13,14,15
NGC 1097	88	$\lesssim 10$	$\sim 10^6$			16
NGC 6951	24	$\lesssim 10$	$\sim 10^6$			16
Arp 220	> 8	$\lesssim 100$:		9×10^9		17,18
NGC 3597	$\simeq 70$	~ 200?	$10^7 - 10^8$:	3×10^9		19,20,21
NGC 7252	$\simeq 140$	700	4×10^7	3.5×10^9		22,23
NGC 3256	hundreds	$\gtrsim 100$	6×10^7	1.5×10^{10}		24

Sources: (1) Holtzman et al. 1992 (2) Carlson et al. 1998 (3) Lazareff et al. 1989 (4) Jaffe 1990 (5) Bridges & Irwin 1998 (6) O'Connell et al. 1995 (7) Lo et al. 1987 (8) Satypal et al. 1997 (9) Watson 1996 (10) Mauersberger et al. 1996 (11) Scoville et al. 1985 (12) Alonso & Minniti 1997 (13) Minniti et al. 1996 (14) Schreier et al. 1996 (15) Eckart et al. 1990 (16) Barth et al. 1995 (17) Scoville 1998 (18) Shaya et al. 1994 (19) Holtzman et al. 1996 (20) Wiklind et al. 1995 (21) Lutz 1991 (22) Miller et al. 1997 (23) Wang et al. 1992 (24) Zepf et al. 1999

These larger galaxies present a much more heterogeneous collection than the simpler starburst dwarfs. Some (NGC 1097, 6951) show star formation along an inner ~ 1 kpc ring of gas; some (M82, NGC 253) may have been stimulated by tidal shocks; some (NGC 1275, 5128) are giant ellipticals which appear to have undergone accretions of smaller gas-rich satellites; and some (NGC 3256, 3597, 7252) are suggested to be merger remnants in the sense used above, i.e. the collisions of two roughly equal disk galaxies. All of them have complex structures and morphologies. For example, the nearby and well studied elliptical NGC 5128 has star clusters in its inner few kpc which appear to be a broad mix of ages and metallicities (some from the original elliptical, some which may have been acquired from the disk-type galaxy it recently accreted, and some bluer objects recently formed out of the accreted gas; see, for example, Alonso & Minniti 1997; Schreier et al. 1996; Minniti et al. 1996). Its halo within $R_{gc} \lesssim 20$ kpc shows the characteristics of a fairly complex triaxial structure (Hui et al. 1995) revealed through the kinematics of both its planetary nebulae and globular clusters.

Nevertheless, these large galaxies display some important features in common with the starburst dwarfs: the massive young star clusters are forming preferentially in the densest, central regions of the collected gas; and their total masses are, once again, of order 1 percent of the residual gas ($HI + H_2$) present in the active regions. It does not appear to matter how the gas is collected; but there needs to be lots of gas collected into SGMC-sized regions before globular clusters can be built.

1.9.7 A Brief Synthesis

In this and Section 8, we have approached the discussion of galaxy formation from a number of different directions (*in situ*, mergers, accretions). How well do these different hypotheses score, as ways to build globular clusters?

By now, it should be apparent that each one of these generic pictures represents an extreme, or limiting, view of the way that galaxies must have assembled. Thus in some sense a "scorecard" is irrelevant: each approach in its extreme form would get a passing grade in some situations but an obvious failing grade in others. In other words, it is becoming increasingly clear that we need elements of *all* these approaches for the complete story. Mergers of disk systems are plainly happening in the present-day universe and should have happened at greater rates in the past. Accretions of small satellites are also happening in front of us, and will continue to take place in the ongoing story of galaxy construction. Yet there must also have been a major element of *in situ* formation, involving the amalgamation of many small gas clouds at early times while the first rounds of star formation were already going on within those pregalactic pieces.

The *in situ* approach imagines that a considerable amount of star formation took place at early times. For large elliptical galaxies, which have dominated much of the discussion in these chapters, there is now much evidence that their main epoch of formation was at redshifts $z \sim 3-5$ (see, e.g., Larson 1990b; Maoz 1990; Turner 1991; Whitmore et al. 1993; Loewenstein & Mushotzky 1996; Mushotzky & Loewenstein 1997; Steidel et al. 1996, 1998; Giavalisco et al. 1996; Bender et al. 1996; Ellis et al. 1997; Stanford et al. 1998; Baugh et al. 1998, for only a few examples of the extensive literature in this area). Evidence is also gathering that ellipticals in sparse groups may also have formed with little delay after the rich-cluster ellipticals (Bernardi et al. 1998; G. Harris et al. 1999), somewhat contrary to the expectations of hierarchical-merging simulations.

Our impression of what a large protogalaxy looked like at early times continues to be influenced strongly by the Searle-Zinn picture: the logical sites of globular cluster formation are $\sim 10^8 - 10^9 M_\odot$ clouds, which are capable of building up the basic power-law mass spectrum of clusters that matches the observations (McLaughlin & Pudritz 1996). We can expect that vigorous star formation should be happening within these SGMCs at the same time as they are combining and spilling together to build up the larger galaxy. For dwarf E galaxies, perhaps a single SGMC or only a few of them combined, and one initial starburst truncated by early mass loss may spell out the main part of the formation history (Section 8 above). For giant E galaxies, dozens or hundreds of SGMCs would have combined (Harris & Pudritz 1994), and there could well have been at least two major epochs of star formation separated by a few Gyr, leaving their traces in the bimodal MDFs of the globular clusters (Forbes et al. 1997) and the halo field stars (G. Harris et al. 1999). The metal-poor part of the halo appears (from preliminary findings in NGC 4472 and NGC 5128, as discussed above) to be remarkably more "cluster-rich" than the metal-rich component, strongly suggesting that a great deal of protogalactic gas ended up being unconverted to stars until the second round or later (McLaughlin 1999); perhaps much of the gas in the original SGMCs was stripped away during infall (Blakeslee 1997) or driven out during the first starburst by galactic winds (Harris et al. 1998a). There are many potential influences to sort out here, and the sequence (and nature) of the events is still murky even without bringing in later influences from mergers and accretions.

The basic *accretion* model starts with a large initial galaxy which had already formed *in situ*, by hypothesis in a single major burst. Around this core, we then add a sequence of smaller galaxies and thus build up the metal-poor halo component. But there are two basic varieties of accretion: gas-free or gas-rich. If the satellites are gas-free dwarf ellipticals, then we should expect to build up a larger galaxy with a specific frequency in the normal range, and a halo MDF that is intermediate or moderately low metallicity. But if the accreted objects are gas-rich, then new clusters and halo stars can form in the process, drive the MDF increasingly toward the metal-rich end, and (if the total amounts of accreted gas are *very* large) possibly change the specific frequency. But in a fairly literal sense, this latter version of building a galaxy by adding together gas-rich dwarfs is quite close to the generic Searle-Zinn picture.

The *merger* model applies specifically for gE galaxy formation. If we amalgamate pre-existing galaxies of roughly equal size, the result should in most cases be an elliptical. But again, the amount of gas will play an important role in the outcome. If the mergers are taking place at high redshift (that is, at very early times), then we can expect the progenitors to be largely gaseous, in which case the majority of stars in the merged product would actually form during the merger of gas clouds. This, too, can be viewed as an extension of the basic Searle-Zinn formation. On the other hand, if the merging is happening at low redshift nearer to the present day, then the galaxies have smaller amounts of gas; much less star formation can happen during the merger, and the result will be a low$-S_N$ elliptical such as we see in small groups.

The presence or absence of gas is therefore a critical factor in evaluating the success of any formation picture. It is only the presence – or the removal – of large amounts of gas during the most active cluster formation epochs which will permit significant changes in the total numbers of globular clusters (and thus the specific frequency), and the form of the metallicity distribution function. The challenge for any one galaxy is to identify the individual combination of events which led to its present-day structure.

Here – appropriately, in the confusion and uncertainty that characterizes the frontier of any active subject – we must end our overview of globular cluster systems. Many areas of this growing subject have been missed, or dealt with insufficiently, and I can only urge readers to explore the rich literature for themselves. The surest prediction is that new surprises await us.

To see the world for a moment as something rich and strange is the private reward of many a discovery.

Edward M. Purcell

ACKNOWLEDGEMENTS

It is a genuine pleasure to thank Lukas Labhardt, Bruno Binggeli, and their colleagues and staff for a superbly organized and enjoyable conference in a beautiful setting. Bruce Carney and Tad Pryor were the ideal fellow lecturers for this concentrated meeting on globular clusters. Finally, I am grateful to Dean McLaughlin and David Hanes, who gave me careful and constructive readings of the manuscript before publication.

1.10 APPENDIX: AN INTRODUCTION TO PHOTOMETRIC MEASUREMENT

Many of the issues in these lectures are strongly related to our ability to *measure* certain characteristics of globular clusters reliably and believably: luminosities, distances, chemical compositions, reddenings, ages .... the potential list is a long one. Many of these tasks boil down to the process of photometric data reduction from CCD images – simple in principle, but surprisingly intricate and challenging in practice.

1.10.1 An Overview of the Color-Magnitude Diagram

To start with, let us look at the essential features of a typical globular star cluster as they appear in the color-magnitude diagram (CMD), Fig. 1.63. Here, all the evolutionary stages of stars in this ancient object are laid out for us to see. First is the zero-age main sequence (ZAMS), where the low-mass stars are quietly undergoing core hydrogen burning (at the bottom end, the luminosity starts to decline steeply as the mass decreases toward the hydrogen-burning limit, and our best attempts at tracing them may be lost in the increasing scatter of photometric measurement uncertainty.) Core hydrogen exhaustion is marked by the turnoff point (MSTO), after which the

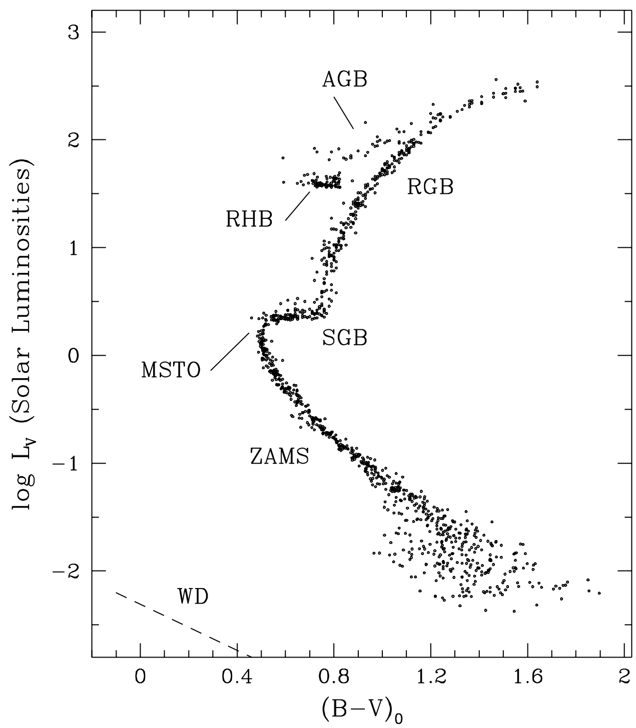

Fig. 1.63. Color-magnitude diagram for a typical globular cluster (data from Hesser et al. 1987, for 47 Tucanae). The axes are plotted as visual luminosity in Solar units, against the ratio of visual to blue luminosity (essentially, the color index $B-V$). The principal evolutionary stages in the stars' history are marked with the abbreviations defined in the text

stars move rapidly across the subgiant branch (SGB), then steadily up the red-giant branch (RGB) as the hydrogen-burning shell gradually moves outward in mass through the stellar interior and the inert helium core gradually increases in mass. Core helium ignition takes place at the RGB tip, and the star rapidly readjusts to a new equilibrium on the horizontal branch (HB). If the cluster has moderately high metallicity or low age, the HB stars will have large hydrogen envelopes and low surface temperatures and will thus all be on the red side of the horizontal branch (RHB). But if the cluster is very old, or has low metallicity, or the stars have suffered high mass loss from their surfaces, then the residual hydrogen envelope will be small, the stars will have high surface temperatures, and the horizontal branch will extend far over to the blue (BHB) side of the CMD.

The last active stage of nuclear burning for globular cluster stars is the asymptotic giant branch (AGB) in which two fusion shell sources (hydrogen to helium, helium to carbon) sit on top of an inert core. When these shells approach too close to the stellar surface, the remaining envelope is ejected as a planetary nebula, the shell-burning sources are permanently extinguished, and the central dead core (a mixture of He/C/O in proportions determined by the original mass, with a tiny surface skin of hydrogen) settles into the white dwarf (WD) phase. The star is now supported mainly by the degeneracy pressure of the electron gas, and as it emits its residual heat, it gradually slides down the WD cooling line to the cold black-dwarf state.

The basic features of the various nuclear burning stages (ZAMS through AGB) have been recognized in the CMD for many years (for comprehensive reviews, see Renzini & Fusi Pecci 1988; Chiosi et al. 1992; and Carney's lectures in this volume). Current evolutionary stellar models and isochrones are now able to reproduce the observed CMDs in considerable detail (see, e.g., Sandquist et al. 1996; Harris et al. 1997b; Salaris et al. 1997; Stetson et al. 1999; Cassisi et al. 1999; Richer et al. 1997; Dorman et al. 1991; Lee et al. 1994, for just a few of the many examples to be found in the literature). However, the extremely faint WD sequence, and the equally faint bottom end of the ZAMS where the masses of the stars approach the hydrogen-burning limit, have come within reach of observation for even the nearest clusters only in very recent years (see Cool et al. 1996; Richer et al. 1997; King et al. 1998, for recent studies that delineate these limits). All in all, the CMDs for globular clusters provide one of the strongest and most comprehensive bodies of evidence that our basic understanding of stellar evolution is on the right track, even if many detailed steps still need work.

1.10.2 Principles of Photometry and The Fundamental Formula

High quality color-magnitude studies are obtained only after carefully designed observations and data reduction.

Fig. 1.64 shows a pair of CCD images obtained with the Hubble Space Telescope WFPC2 cameras. The first is a single exposure with the PC1, after preprocessing.[12] It illustrates a number of problems that we need to attack, once we have our CCD image in hand:

- We need to eliminate *artifacts* from the image (bad pixels, cosmic rays, and so on).
- We need to *find* the real objects (stars, galaxies, asteroids, ...) in an objective and reproducible way.
- We need to *classify* the objects, i.e., divide our objects into separate lists of stars, galaxies, or other things.
- We need to *measure* the brightness and location of each object.

[12] The steps involved in preprocessing — bias subtraction and trim, dark current subtraction, flat fielding – are not discussed here. More detailed discussions of these operations can be found, e.g., in Walker (1987); Tyson (1989); Massey & Jacoby (1992); Howell (1992); Gullixson (1992); or Gilliland (1992).

- We need to understand the *uncertainties* in the measurement, both random and systematic.

Eliminating unwanted artifacts from a single raw image such as that in Figure 1.64 can be done, to some extent, with sophisticated rejection algorithms which look for sharp features (hot or very cold pixels, or cosmic-ray hits which affect only one or two pixels); or elongated or asymmetric features (bad columns, cosmic-ray streaks, bright stars which bleed along rows or columns). You can then attempt to smooth over these places with the hope that the adjacent pixels will allow you to reconstruct the information on these damaged pixels correctly. This approach will not work so well if the star images themselves are undersampled, such as in the WF frames of the HST, or if the artifact happens to be embedded somewhere in a real object (a star or galaxy). A much more effective process in any case is to take a series of images that are deliberately shifted in position betwen exposures (sub-pixel-shifted or "dithered") so that the pixel grid samples the field at several different positions.[13] When the frames are re-registered, and combined through a median or averaging algorithm which rejects extreme values, the artifacts will almost entirely drop out, leaving an enormously cleaner combined frame. Obviously, the more frames you can use to do this, the better. If several independent positions have been sampled differing by fractions of a pixel, a higher-resolution summed image with finer pixels can also be constructed (see, e.g., the 'drizzle' software of Hook & Fruchter 1997 and Fruchter & Hook 1999, or the ALLFRAME code of Stetson 1994), yielding gains in the ability to measure crowded objects or to determine image structure. Image reconstruction from images that have been carefully sub-pixel-shifted can yield powerful improvements especially in image structure or morphology studies (see Lauer 1999 for a review of reconstruction algorithms). However, reconstructed images may not be suitable for photometry since some algorithms do not conserve flux; be cautious in making a choice of techniques.

Photometry of *stars* is a far easier job than photometry of nonstellar objects, for one predominant reason: on a given CCD image, all stars – which are point sources of light blurred by the atmosphere (for ground-based telescopes) and the telescope optics – have the same profile shape called the point spread function (PSF).[14] The PSF width can be characterized roughly by the full width at half-maximum height (FWHM) of the profile.

[13] "Dither" literally means "to act nervously or indecisively; vacillate". This is an unfortunate choice of term to describe a strategy of sub-pixel-shifting which is quite deliberate!

[14] Strictly speaking, the assumption that all stars have the same PSF on a given image is only true to first order, since image scale or optical aberrations can differ subtly across the field of view in even the best situations. An important example is in the HST/WFPC2 cameras, but even here the variations can be fairly simply characterized; see the WFPC2 Instrument Handbook available through the StScI website. More generally, we can say that the PSF for star images should be a *known and slowly changing function of position on the frame*, and should be

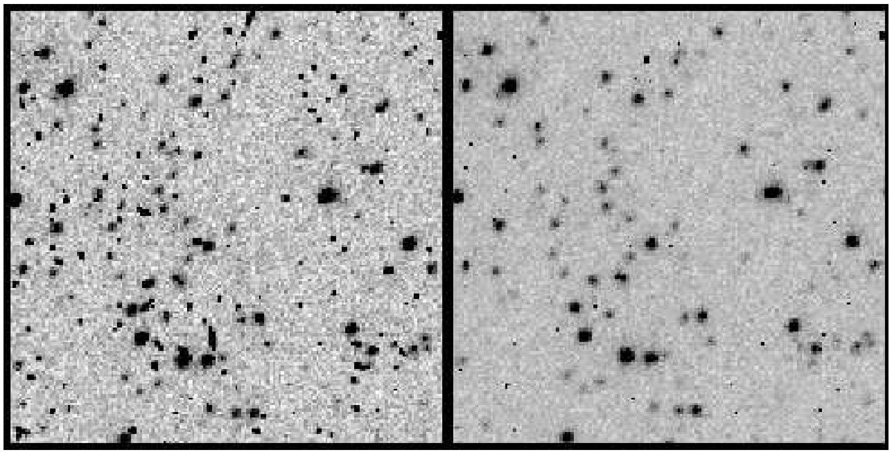

Fig. 1.64. *Left panel:* A segment of a single exposure in the field of globular cluster NGC 2419, taken with the HST (PC1 camera). Many star images are present, but the frame is heavily contaminated by numerous cosmic-ray hits and bad pixels. *Right panel:* The same field after re-registering and combining 8 individual exposures that were sub-pixel-shifted. Only the stars and a few easily distinguished bad pixels remain, on a much smoother background. Raw image data are taken from the study of Harris et al. (1997)

Finding stars in an objective manner is a straightforward job in principle: look for objects whose brightest pixels stand clearly above the pixel-to-pixel scatter of the sky background (see Fig. 1.65). If we define z_s as the mean sky brightness, and σ_s as the standard deviation of the sky pixels, then we can set a *detection threshold* for real objects by looking for any pixels with intensities $z > (z_s + k \cdot \sigma_s)$ for some threshold parameter k. Thresholds in the range $k \simeq 3.5 - 4.0$ are normal for faint stellar photometry; one can try to go lower, but choices $k \lesssim 3$ inevitably lead to lots of false detections and serious contamination problems (see below).

The sky characteristics (z_s, σ_s) are *local* quantities which may differ strongly across the image field. The all-important noise parameter σ_s, which directly fixes the faint limit of your photometry, is governed by (a) the raw sky brightness z_s through simple photon statistics (are you working in the wings of a large galaxy or nebula which covers much of the frame?); (b) the 'lumpiness' of the sky (are you working in a crowded field, or trying to find stars within a patchy nebula or a spiral arm of a distant galaxy?); (c) bad pixels and cosmic rays, if you were unable to remove those; (d) instrumental noise such as readout noise and dark current; (e) additional noise introduced in the pre-

independent of brightness as long as the detector is linear. Neither of these statements will be true for nonstellar objects such as galaxies, whose profile shapes cover an enormously broader parameter space.

processing, such as flat-fielding (did your flat-field exposures have inadequate signal in them compared with the target exposures?).

The principles of simple aperture photometry will illustrate several of the basic limits inherent in photometric measurement. Suppose that we have a star located on a grid of pixels, as in Fig. 1.66, and that its intensity profile is given by the matrix $z(x, y)$. As before, denote the local sky and background noise as (z_s, σ_s). Suppose we now surround the star with a circular measuring aperture of radius r, which will then contain n_{px} pixels approximately given by $n_{px} \simeq \pi r^2$. (The finite pixellation of the image means that the ideal circular aperture boundary becomes a jagged line enclosing only whole pixels. In some photometry codes including DAOPHOT, second-order corrections are made to round off the boundary by adding fractional amounts of light from pixels along the rim of the circle.) Now let

$N_\star$ = number of collected e^- from the star image

N_s = number of collected e^- from the sky background light in the aperture

I = variance per pixel (in e^-) of the instrumental noise (including readout noise, dark current, quantization noise, and perhaps other factors).

These three quantities are normally uncorrelated, so we can add their variances to obtain the total variance of the random noise in the aperture. The *signal-to-noise ratio* of the measurement is then

$$\frac{S}{N} = \frac{N_\star}{\sqrt{N_\star + N_s + n_{px} I}} \tag{1.66}$$

This ratio represents the internal uncertainty in the measured brightness of the star. The *instrumental magnitude* of the star (for example, in the V filter) is normally expressed as

$$v_{inst} \equiv -2.5 \log\,(N_\star/t)\;+\;\text{const}$$

where t is the exposure time and thus $N_\star/t \equiv r_\star$ is the "count rate" (e^- per second) from the star. The last term is an arbitrary constant. The instrumental magnitudes can be transformed to standard V magnitudes by measurements of photometric standard stars with known magnitudes and colors, taken during the same sequence of observations as the program exposures. For a thorough outline of precepts for standard-star transformations, see Harris et al. (1981).

What governs each of the factors in the uncertainty?

- $N_\star$ is proportional to the photon collection rate $r_\star$ from the star; the exposure time t; and the detection efficiency Q (conversion ratio of incident photons to stored electrons, characteristic of the detector). In turn, $r_\star$ is proportional to the brightness of the star $b_\star$ and the telescope aperture size D^2 (the collecting area).

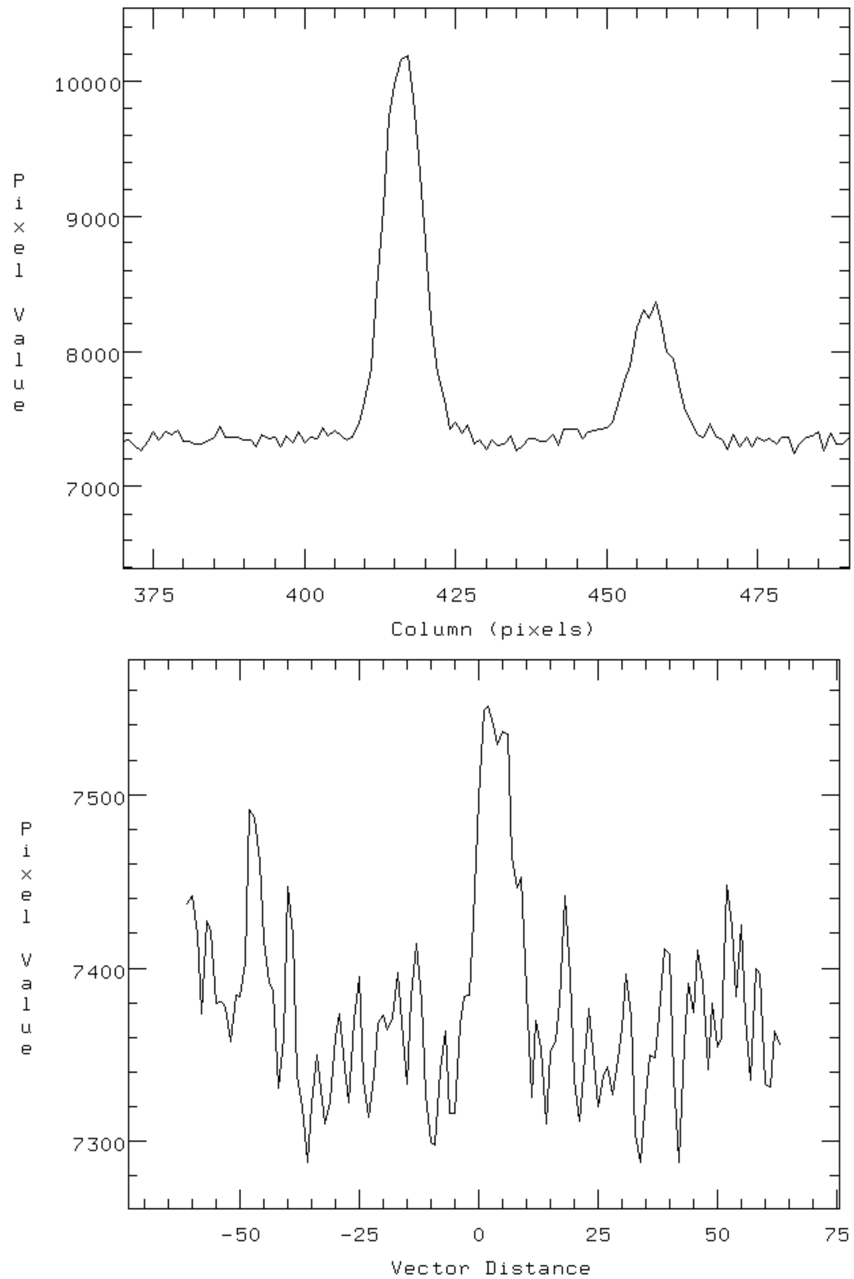

Fig. 1.65. **(a)** Profile for two bright stars on a CCD frame; these stand clearly above the sky background noise. **(b)** Profile for a much fainter star on the same frame. This star is near the limit of detectability relative to the standard deviation of the sky background noise

- N_s is proportional to t, Q, D^2, n_{px}, and μ_s where μ_s is the sky brightness (number of photons per second per unit area). Also, we have $n_{px} \sim r^2 \sim \alpha^2$ where α denotes the star image size (FWHM). Smaller seeing disks can be surrounded by smaller apertures, thus lowering the amount of contaminating skylight and improving the signal-to-noise.
- I (instrumental noise per pixel) is approximately constant if dark current is negligible; if dark current dominates, however, then $I \sim t$.

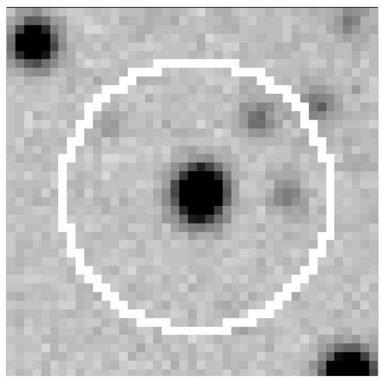

Fig. 1.66. CCD image of a bright star, surrounded by a digital 'aperture' (white line). The aperture boundary follows the quantized pixels and is only an approximation to a circle. Note three faint stars inside the aperture, which will contaminate the measurement of the central bright star

For most broad-band imaging applications and modern low-noise CCD detectors (which characteristically have $I \lesssim 5e^-/\mathrm{px}$), the instrumental noise is not important. (A notable exception is, again, HST/WFPC2 where the readnoise $n_{px}I$ is roughly equal to the sky noise N_s even for full-orbit exposures and broad-band filters.) In addition, we are usually interested in how well we can do at the faint limit where $N_\star \ll N_s$. So the most interesting limit of the S/N formula is the "sky-limited" case,

$$\frac{S}{N} \simeq \frac{N_\star}{\sqrt{N_s}} \,. \tag{1.67}$$

If we put in our scaling laws $N_\star \sim b_\star \, t \, Q \, D^2$ and $N_s \sim \mu_s \, t \, Q \, D^2 \alpha^2$, we obtain

$$\left(\frac{S}{N}\right)^2 \sim \frac{b_\star^2 \, t \, Q \, D^2}{\alpha^2 \mu_s} \,. \tag{1.68}$$

Clearly there are several ways we can achieve deeper photometric limits. We can:

- Increase the exposure time.
- Use a bigger telescope.
- Find darker sky.
- Improve the seeing quality.
- Employ a better detector (lower noise or higher Q).

The circumstances will determine which routes are possible at any given time. Modern optical CCD detectors have detection efficiencies Q approaching 1, and readnoise levels I of only a few e^-, so they can be "improved" substantially only by making them in physically larger arrays or extending their wavelength coverage. Much effort has been directed toward techniques for improving image quality (seeing), either by space-based observations or by adaptive optics from the ground. It is important to note that better seeing yields other gains beyond the formal improvement in S/N: a narrower stellar profile will also reduce contamination from image crowding, and allow us to see more detailed structure within nonstellar objects.

The gains in limiting magnitude in astronomical photometry over the past two decades have been spectacular (see Fig. 1.67): more than an order of magnitude of depth was gained in the early 1980's with the appearance of CCD detectors, which were ~ 100 times higher in quantum efficiency than photographic plates. Another decade later, *HST* was able to reach another order of magnitude deeper because of the jump in spatial resolution (FWHM $\simeq 0.''1$ as opposed to $\gtrsim 0.''5$ from the ground) and the considerably darker sky as observed from space. Still further gains in the post-*HST* era will require larger apertures and longer exposure times.

A useful illustration of the way the factors in (1.68) complement each other is to compare two similar photometric experiments carried out in completely different eras. Recently, Harris et al. (1998b) used the *HST* cameras to resolve the brightest old-halo RGB stars in a dwarf elliptical galaxy in the Virgo cluster, at a distance of $d \simeq 16$ Mpc (see Fig. 1.68). Half a century earlier, Walter Baade (1944) achieved exactly the same thing for the Local Group dwarf ellipticals NGC 185 and 205 ($d = 0.8$ Mpc), in a classic experiment which first showed that these galaxies were built of old-halo stars and were companions to M31. Baade used a total exposure time of 4 hours with the then-new red-sensitized photographic plates, on the Mount Wilson 2.5-meter telescope. The seeing was probably $\alpha \simeq 1''$; the efficiency Q of the emulsion is hard to guess but would certainly have been less than 1%. By comparison, the *HST* is a 2.4-meter telescope, the DQE of WFPC2 is $Q \simeq 30\%$, the "seeing" is $\alpha = 0.''1$, and the total exposure on the Virgo dwarf was 9 hours. Both experiments measured stars of the same absolute magnitude, so the relative photon collection rate scales just as the inverse square of the distance, $b_\star \sim d^{-2}$. Putting these factors into (1.68), we have as our comparison for the two experiments

$$\frac{\frac{S}{N}(Baade)}{\frac{S}{N}(HST)} \simeq \frac{\left(\frac{4}{9} \cdot \frac{1}{50}\right)^{1/2} \cdot 1}{\left(\frac{1}{20}\right)^2 \cdot 10 \cdot 10^{1/2}} \simeq 1.2\,.$$

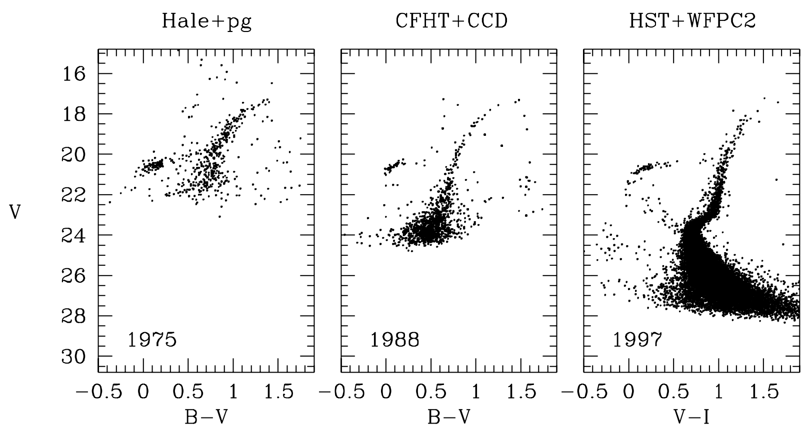

Fig. 1.67. Three color-magnitude diagrams for the remote halo cluster NGC 2419, showing the gain in depth from increased DQE and spatial resolution. *Left panel*: CMD containing 547 stars (Racine & Harris 1975), obtained via photographic plates with the Hale 5-m telescope. *Center panel*: CMD containing 1316 stars (Christian & Heasley 1988), obtained with CCD photometry from the CFHT. *Right panel*: CMD containing 17275 stars (Harris et al. 1997b), obtained with the WFPC2 CCD camera on the Hubble Space Telescope. The total exposure times in all three cases were similar

The photometric limits of both experiments turn out to be similar – both of them reach just about one magnitude below the RGB tip of their respective target galaxies, just as the quantitative comparison would predict. The distance ratio is obviously the biggest factor in the equation; it makes the same types of stars in the Virgo dwarf 400 times fainter in apparent brightness than those in the Local Group ellipticals.

1.10.3 Aperture and PSF Measurement

We have already discussed the initial process of determining the sky background noise σ_s and finding stars across the field. Now we need to decide how to measure them. Most of all, this decision depends on whether the frame is *crowded* or *uncrowded*. Suppose the picture has a total area of A pixels and there are n detected objects, so that the average area per object is $\simeq A/n$ px. A useful guideline is that if $(A/n) \gtrsim \pi\,(4\,FWHM)^2 \simeq 50\,(FWHM)^2$, then the image is not very crowded in absolute terms.

For **uncrowded images**, straightforward aperture photometry (Fig. 1.66) will work well, and is conceptually the simplest measurement technique. To understand the empirical image profile, it will help to take several bright stars and plot up the *curve of growth* – i.e., the apparent magnitude of the star as a function of aperture radius – to find out how large an aperture one needs to enclose virtually all the light of the star. Choose a radius r_{max}

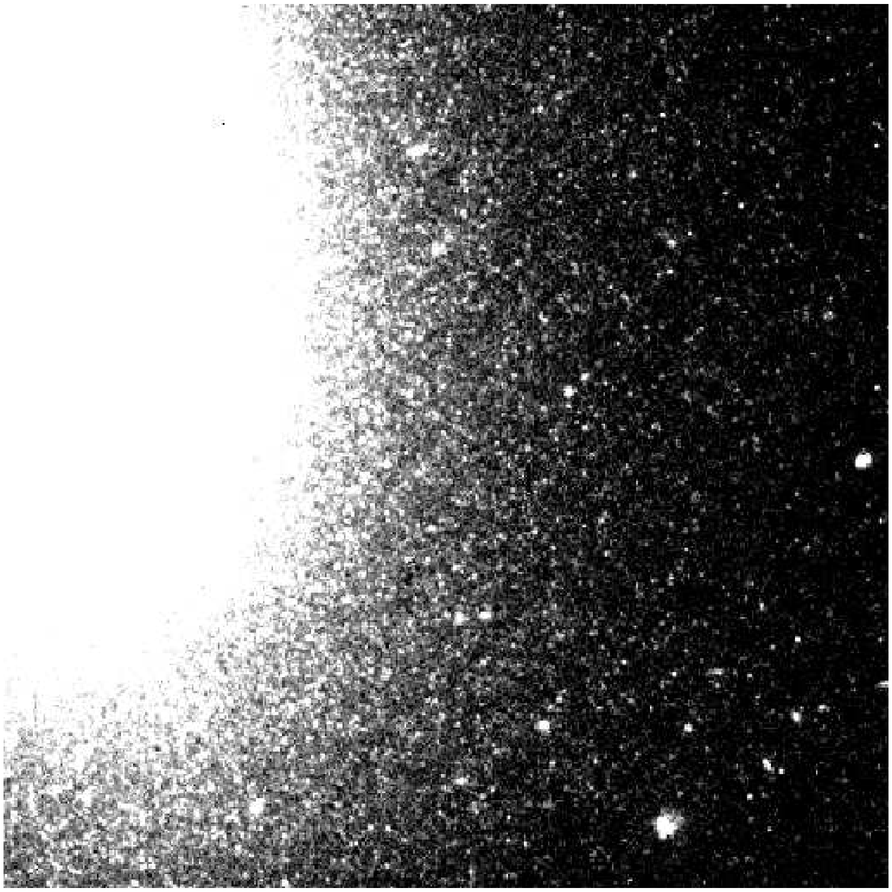

Fig. 1.68. *HST* image in the $I-$band of the Virgo dwarf elliptical VCC 1104, from Harris et al. (1998b). Almost one magnitude of the galaxy's old red giant branch is clearly resolved

which contains, e.g., $\gtrsim 97\%$ of the asymptotic total, but which is not so large that sky noise starts to affect the internal uncertainty. This is the best aperture size to use for calibration purposes, since it will be unaffected by minor differences in seeing (which affects the core image structure) from place to place on the frame, or between frames taken at different times during the same observing session. We measure known standard stars on other frames through the same aperture and thus define the transformation between the "instrumental" magnitude scale v into true magnitude V.

However, to compare the relative magnitudes of all the stars on one frame, both bright and faint, we should instead choose an aperture radius which maximizes S/N. For the bright stars where $N_\star \gg N_s$, we have $(S/N) \simeq \sqrt{N_\star}$ and a large aperture is best. But the vast majority of the stars on the frame

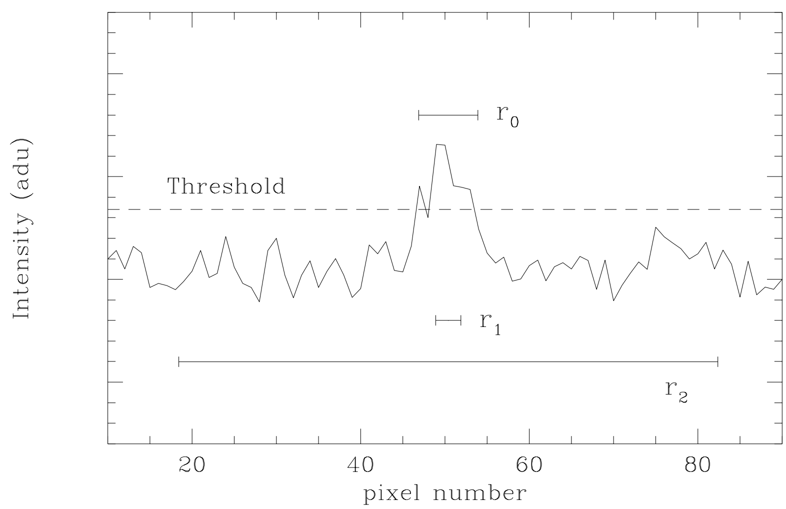

Fig. 1.69. Definition of the optimum radius for aperture photometry. A faint-star image profile superimposed on sky noise is shown, with its central few pixels extending up past the detection threshold (dashed line). A small aperture (r_1) contains too small a fraction of the star light to produce adequate S/N, while a large one (r_2) is dominated by sky noise. The intermediate aperture r_0 maximizes the signal-to-noise ratio

will be faint ($N_\star \ll N_s$), and for them we have $(S/N) \simeq N_\star/\sqrt{N_s}$. Suppose the star intensity profile is given by $I(r)$. Then

$$N_\star = \int_0^r I(\rho) \cdot 2\pi\,\rho\,d\rho \tag{1.69}$$

and thus the signal-to-noise scales as

$$\frac{S}{N} \sim \frac{1}{r} \int \rho\, I\, d\rho\,. \tag{1.70}$$

Once we know the profile shape $I(r)$, we can solve (1.70) numerically to find the optimum radius r_0. For a Gaussian profile, we have $I(r) = I_0 \exp(-r^2/2\sigma^2)$ and the maximum S/N occurs for an intermediate radius $r_0 \simeq 1.6\,\sigma \simeq 0.67$ FWHM (see Pritchet & Kline 1981). Essentially, this optimum radius is big enough to include most of the starlight, but not so big that a large amount of sky background intrudes to dominate the scatter. The point is illustrated in Figure 1.69.

Once you have all the magnitudes measured through the optimum radius r_0, the brighter stars can be used to find the mean correction from r_0 to r_{max}, so that the whole list is then calibrated.

For **crowded images**, we have to plunge all the way into the more formidable job of profile fitting. One good approach is to use the bright stars with high S/N to define the PSF shape empirically. (Pure analytic approximations to the PSF, such as Gaussians or Moffat profiles, can also be effective and may work in situations where empirical PSFs are difficult to derive.) If the shape parameters (FWHM, noncircularity, orientation) depend on location (x, y) in the frame, then typically a few dozen stars spread evenly over the frame will be needed to map it out; the more the better. Once you have defined the PSF, then try to fit it to each object in the detection list. For *each* star there are at least four adjustable parameters: the object center (x_c, y_c), the brightness scale factor A, and the predicted local sky level $z_s(x_c, y_c)$ at the object center. To test the quality of the result, subtract off the fitted PSF at each star and look for anomalous or distorted residuals. In practice, any one star image will overlap to varying degrees with the wings of all its neighbors, so that the full-blown solution actually requires a simultaneous and highly nonlinear fit of the PSF model to all the stars at once.

The fitted quantity we are most interested in is the scale factor A (unless we are doing astrometry!). Fortunately, the exact form of the model PSF does not have a major effect on A, since both faint and bright stars have the same profile shape and we only need to know their relative brightness scale factors. A more accurate PSF will, however, allow you to do a better job of subtracting out neighboring stars that are crowding your target star, so it is worth spending time to get the best possible PSF. For very crowded fields, defining the PSF itself is an iterative and sometimes painful business: one must make a first rough PSF, use that to subtract out the neighboring objects around the stars that defined the PSF, then get a cleaner PSF from them and repeat the steps. Finding stars, too, is an iterative process; after the first pass of PSF fitting and subtraction, additional faint stars are often found that were hiding in the wings of the brighter ones or somehow lost in the first pass. These should be added to the starlist and the solution repeated.

The ability to fit *and subtract* stars from the frame is one of the most powerful and helpful features of digital photometry. For example, it can be used even to improve simple aperture photometry, by "cleaning" the area around each measured star even if it is not severely crowded in absolute terms (see, e.g., Stetson 1990; Cool et al. 1996). Nevertheless, in extreme situations the practice of PSF-fitting photometry may take a great deal of the photometrist's time and thought. The best work is still something of an art.

All the basic steps discussed above can be turned into automated algorithms, and several flexible and powerful codes of this type are available in the literature (DAOPHOT, DoPHOT, Romafot, and others; see Stetson 1987, 1994; Stetson et al. 1990; Schechter et al. 1993; Buonanno & Iannicola 1989; Mighell & Rich 1995). These papers, as well as other sources such as Stetson's DAOPHOT manual, supply more advanced discussions of the process of CCD stellar photometry.

1.10.4 Testing the Data

If stars can be subtracted from the frame, they can also be added. That is, we can put artificial stars (scaled PSFs) into the image at arbitrarily chosen brightnesses and locations, and then detect and measure these simulated objects in the same way that we did the real stars. Since these added stars are built from the actual PSF, and are put onto the real sky background, they resemble the real stars quite closely. The huge advantage is that we know beforehand just how bright they are and exactly where we put them. The ability to create simulated images that resemble the real ones in almost every respect is a powerful way to test the data and understand quantitatively our measurement uncertainties and systematic errors.

Three extremely important results emerge from the analysis of simulated images:

- We can determine the *completeness of detection* $f(m)$: that is, we can find out what fraction of the artificially added stars were picked up in the normal object-finding process, as a function of magnitude m.
- We can determine any *systematic bias* $\Delta(m)$ in the measured magnitudes, again as a function of magnitude.
- We can determine the *random uncertainty* $\sigma(m)$ in the measured magnitudes. Here $\sigma(m)$, in magnitude units, is related to our earlier S/N ratio approximately by
$$\sigma \simeq 2.5 \log \left(1 + \frac{1}{(S/N)} \right) .$$

The bright stars are easy to deal with: they will almost all be found ($f \simeq 1$), and measured without bias ($\Delta \simeq 0$) and with low random uncertainty ($\sigma \to 0$). The real problems show up at the opposite end of the scale:

First, at progressively fainter levels, more and more stars fall below the threshold of detection and the completeness fraction f becomes small. At the 50% completeness level, the brightest pixel in the object is nominally *just* at the detection threshold, so it has an equal chance of falling above or below it depending on photon statistics from both the star and the sky it is sitting on. Fainter stars would nominally never have a bright enough pixel to sit above threshold, but they will be found (though with lower probability) if they happen to fall on a brighter than average patch of sky pixels. Similarly, a star just a little brighter than the nominal threshold can be missed if its local sky level is lower than average. Thus in practice, the transition from $f \simeq 1$ to $f \simeq 0$ is a smooth declining curve (see Fig. 1.70). A simple analytic function due to C. Pritchet which accurately matches most real $f(m)$ curves is

$$f = \frac{1}{2}\left(1 - \frac{a(m-m_0)}{\sqrt{1+a^2(m-m_0)^2}}\right) \tag{1.71}$$

where the two free parameters are m_0 (the magnitude at which f is exactly $1/2$) and a (which governs the slope; higher values of a correspond to steeper downturns). Completeness also depends significantly on the degree of crowding, and the intensity level of the sky; that is, it is a lot harder to find objects in extreme crowding conditions, or on a noisier background.

Second, the systematic bias $\Delta(m)$ starts to grow dramatically at fainter magnitudes, primarily because the fainter stars will be found more easily if they are sitting on brighter patches of background, which produces a measured magnitude brighter than it should be. See Fig. 1.71, and notice that $\Delta(m)$ rises *exponentially* as we go fainter than the 50% completeness level.

Third, the random uncertainty in the measured magnitude increases at fainter levels as the signal declines toward zero and the surrounding noise dominates more and more. Under a wide range of practical conditions, it can be shown that the 50% completeness level corresponds to a typical uncertainty $\sigma \sim 0.2$ mag or $S/N \sim 5$ (see Harris 1990a).

Examples of all three of these effects are illustrated in Figs. 1.70, 1.71, and 1.72. The messages from these simulation studies are clear. All aspects of your data will become seriously unreliable below the crossing point of $\sim 50\%$ completeness, and you need to know where that point is. Do not be tempted to believe any interesting features of your data that you think you see at still fainter levels; and above all, do not publish them!

1.10.5 Dealing With Nonstellar Objects

In most photometry projects, it is extremely helpful to be able to separate out starlike objects from nonstellar things. The former will include true stars, and also other objects that you may be trying to find in distant galaxies such as faint globular clusters, HII regions, galactic nuclei, and so on. The latter category will include anything that does not match the stellar PSF: small, faint background galaxies, resolved nebulae and clusters, unresolved clumps of stars, or even artifacts on the image.

A variety of algorithms can be constructed to *classify* the objects that you find on the image. The rather restricted question we ask during the process of stellar photometry is: how well does the PSF fit a given object? We now replace it by a subtly different and more general one: what parameters can we construct to *maximize the difference between starlike and nonstellar objects*?. And since the range of parameter space occupied by nonstellar objects is much larger than for starlike ones, the "best" answer may depend on the situation.

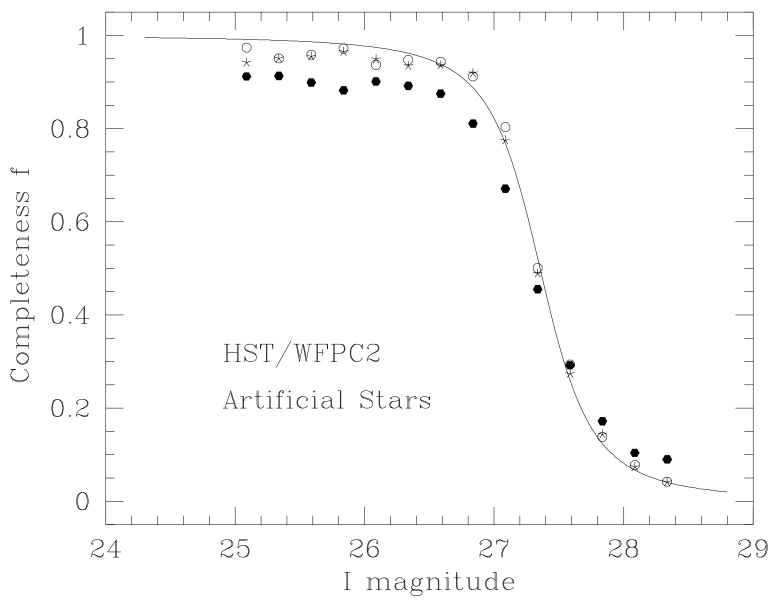

Fig. 1.70. Completeness of detection f, plotted as a function of magnitude, for images taken with the HST WFPC2 camera. Data are from the $I-$band photometry of the Virgo dwarf elliptical shown in Fig. 1.68. The different symbols represent three different regions of the WF2 frame: open and starred symbols are from relatively uncrowded areas, whereas the solid dots are from areas closest to the galaxy center and thus most affected by crowding. The plotted points are binned means of several thousand artificial stars spread over all magnitudes. The model line is the Pritchet interpolation function defined in the text, with parameters $m_0 = 27.36$ and $a = 2.37$. Typically, f declines smoothly from nearly unity to nearly zero over roughly a one-magnitude run of image brightness

The answer to this question generally depends on using the fact that nonstellar objects have more extended contours and radial shapes than the starlike objects on the image. To quantify the characteristic "size" or extent of the object, let us define a general *radial image moment* which is constructed from the pixels within the object (e.g., Tyson & Jarvis 1979; Kron 1980; Harris et al. 1991):

$$r_n = \left(\frac{\sum r_i^n \cdot z_i}{\sum z_i} \right)^{1/n} \tag{1.72}$$

Here z_i is the intensity of the ith pixel above sky background; r_i is the radial distance of each pixel from the center of the object; and there are $i = (1, \ldots, N)$ pixels in the image brighter than some chosen threshold. The object can have any arbitrary shape, so that the sum is simply taken over

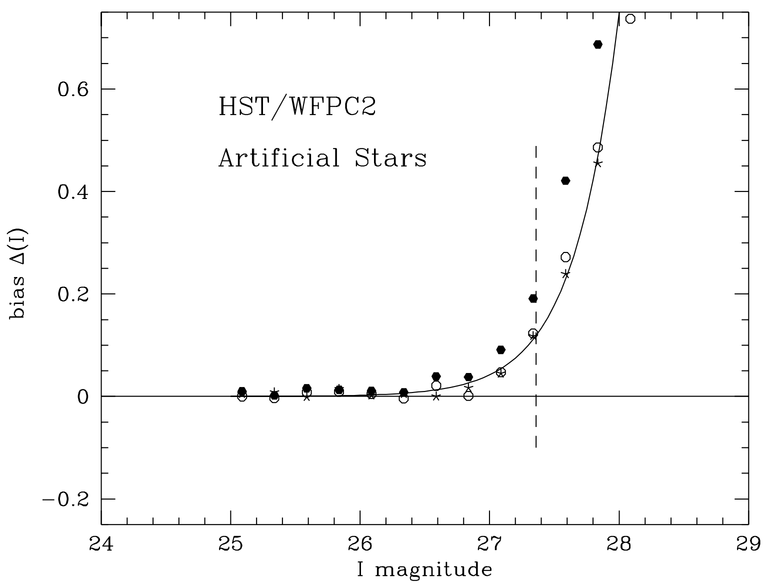

Fig. 1.71. Systematic measurement bias, defined as $\Delta = m(\text{input}) - m(\text{measured})$, plotted against magnitude. Data are from the same sample as in the previous figure. The 50% detection completeness level is marked with the dashed line, while the solid curve shows an exponential function in magnitude fitted to the data points

all the connected pixels making up the object rather than within any fixed aperture. Clearly, r_n represents a *characteristic radius* for the object in pixel units, calculated from the nth radial moment of the intensity distribution. Nonstellar objects (as well as some random clumps of noise) will have larger wings and asymmetric shapes, and thus have larger r_n values than stars do at the same total brightness. A simple plot of r_n against magnitude then effectively separates out the stars from other types of objects.

The choice of weighting exponent n is usually not crucial: positive $n-$values will weight the more extended wings of nonstellar objects more highly, while negative n gives more weight to the sharper cores of starlike objects. The simplest nontrivial moment, r_1, turns out in most situations to work well at separating out a high fraction of the distinguishable nonstellar images (cf. Harris et al. 1991). Other types of moments can be constructed that are sensitive to (for example) image asymmetry, or are linear combinations of various radial and nonradial moments (e.g., Jarvis & Tyson 1981; Valdes et al. 1983).

Though these quantities can be effective at picking out nonstellar objects, they are also good at finding crowded stars which one may not want to elim-

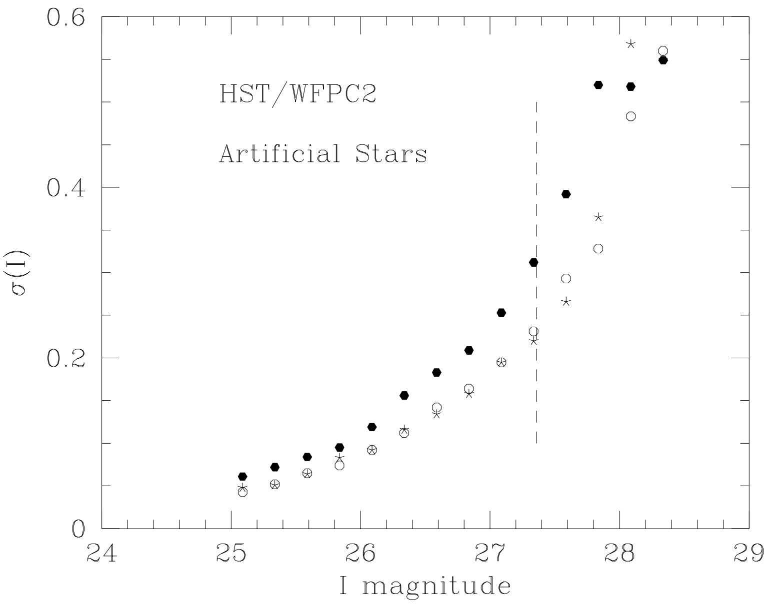

Fig. 1.72. Random measurement error $\sigma(m)$, defined as the root mean square scatter of the magnitude differences (input − measured). Data are from the same sample as in the previous figures. The 50% detection completeness level is marked with the dashed line

inate. Choose what works best for your situation. In some cases, it is important to distinguish between *crowded* pairs of objects (star/star, star/galaxy, galaxy/galaxy) and *single* galaxies with complex, lumpy structures; developing an unbiased algorithm to separate out these cases is a nontrivial exercise (e.g., Jarvis & Tyson 1981; Bertin & Arnouts 1996).

A simple example of one of these diagnostic graphs is shown in Fig. 1.73. The datapoints in the figure represent a mixture of small nonstellar objects (the points falling above the dashed line, which are mostly faint background galaxies) and starlike objects (the points below the dashed line, which are mostly foreground stars and globular clusters belonging to the target galaxy IC 4051). Clearly, the great majority of the nonstellar objects can be separated out cleanly. (Note in this example that all the starlike objects do not have the same characteristic "size" r_1; instead, r_1 decreases for fainter objects. This is because the moment sum defining r_n is taken only over all pixels above a certain threshold, so that the fainter objects have fewer included pixels and smaller characteristic moments. For this reason, the points at the faint end (lower right corner of the figure) fall rather noticeably into quantized groups which represent the small numbers of pixels defining the

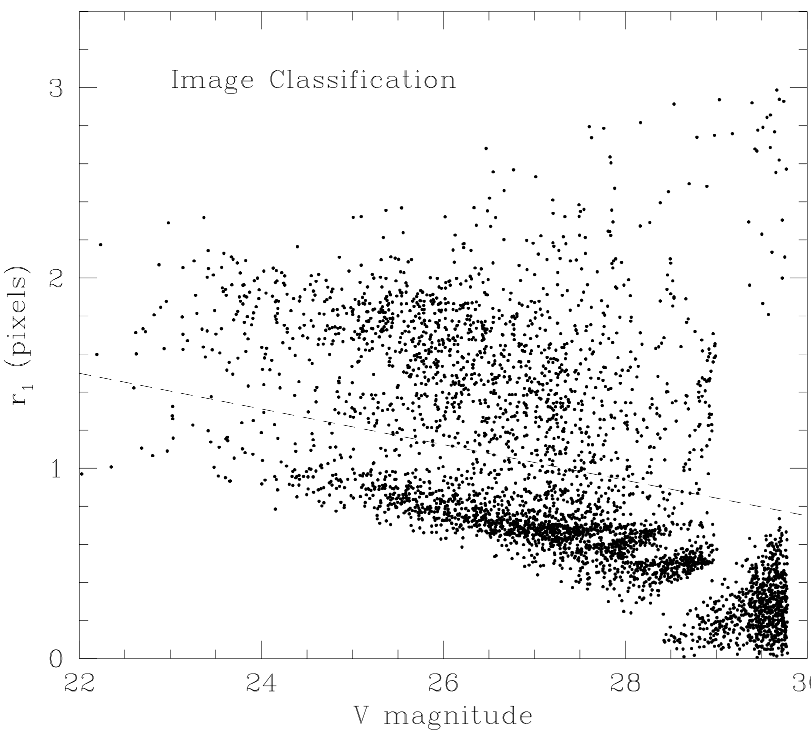

Fig. 1.73. Image structure diagnostic graph for a sample of measurements taken from deep HST/WFPC2 exposures of the remote elliptical galaxy IC 4051 (adapted from Woodworth & Harris 1999). The radial image moment r_1 defined in (1.72) is plotted against V magnitude for $\simeq$ 4500 objects measured on the WF2,3,4 CCD fields

intensity sum. Also, the V magnitude against which the radial moment is plotted is actually not a true fixed-aperture magnitude, but is instead something resembling an "isophotal" magnitude defined from the intensity sum of all pixels brighter than the threshold used in the calculation of r_1. That is, we have $V = \text{const} -2.5 \log (\sum z_i)$. See Harris et al. (1991) for additional discussion.)

It should be stressed that the image moment quantities described here are used *only* for classification purposes and not for actual measurement of the total magnitudes and colors (see below). Thus, they can be defined in whatever way will maximize the difference between stellar and nonstellar objects. Once they have served their purpose of separating out the two kinds of objects, they can be put aside.

The last stage of the classification process is to draw appropriate boundary lines between stellar and nonstellar regions of your chosen diagnostic diagram, and extract the unwanted ones from your object lists. In the simple case defined above, we would use the single diagram of r_1 versus total magnitude and define one empirical boundary line (shown in Fig. 1.73). But in principle, we could simultaneously use many more parameters, such as the aperture growth curve, peak intensity, and nonradial moments. Bertin & Arnouts (1996) nicely describe this step as mapping out the *frontier* between stellar and nonstellar objects in the multi-dimensional parameter space. Where to define the boundary of the frontier is always a matter of judgement, and artificial-star tests can be extremely helpful here for deciding where to place it. The codes of Valdes et al. (1983) and Bertin & Arnouts (1996) combine several image parameters in a manner equivalent to a neural network, and employ simulated images as training sets for the neural net.

At the faint end of the photometry, random noise eventually overwhelms the ability of even the most advanced decision-making algorithms to discriminate between stellar and nonstellar objects. Nevertheless, image classification is worth doing. The eventual payoff is that the contamination "noise" in your selected sample of objects can be tremendously reduced, and in some cases it is critical to the ability to define the sample at all.

In addition, the noise for one experiment can literally be the signal for another: for example, in a deep high-latitude field we might want to study the population of faint galaxies, in which case the foreground stars would be the contaminants. But if we want to study the Milky Way halo stars, exactly the opposite is true.

Measuring the *total magnitudes* of nonstellar objects correctly and consistently is a nontrivial job with a whole new set of special problems: unlike stellar images, there is no "PSF" to refer to, and the parameter space of image properties is vastly larger. One defensible and widely applicable approach is to measure the total magnitude within an aperture which encloses some large fraction (say 90%) of the object's asymptotic total flux, while not becoming so large that the enclosed light is too sensitive to sky noise. Since no two nonstellar objects have the same shape, this optimum aperture will have a different numerical value for each object and must be determined for each one. Under fairly general conditions, the first-order radial moment r_1 defined above is nearly equal to the half-light radius of a faint galaxy; a radius $2r_1$ turns out to enclose $\simeq 90\%$ of the total light and gives a reasonable choice for the optimum aperture magnitude. For detailed discussions of this method, see, for example, Kron (1980); Infante (1987); Infante & Pritchet (1992); Bershady et al. (1994); Bertin & Arnouts (1996); or Secker & Harris (1997), among others. Readers can refer to these same papers for an entry to the extensive literature on this subject.

References

1. Abraham, R. G., & van den Bergh, S. 1995, ApJ, 438, 218
2. Ajhar, E. A., Blakeslee, J. P., & Tonry, J. L. 1994, AJ, 108, 2087
3. Ajhar, E. A. et al. 1996, AJ, 111, 1110
4. Alcock, C. et al. 1997, ApJ, 482, 89
5. Allen, S. W., & Fabian, A. C. 1997, MNRAS, 286, 583
6. Alonso, M V., & Minniti, D. 1997, ApJS, 109, 397
7. Armandroff, T. E. 1989, AJ, 97, 375
8. Armandroff, T. E. 1993, in The Globular Cluster – Galaxy Connection, ASP Conf. Series 48, edited by G. Smith & J. P. Brodie (San Francisco: ASP), 48
9. Armandroff, T. E., Olszewski, E. W., & Pryor, C. 1995, AJ, 110, 2131
10. Arp, H. C. 1965, in Galactic Structure, ed. A. Blaauw & M. Schmidt (Chicago: U. Chicago Press), 401
11. Ashman, K. M., Bird, C. M., & Zepf, S. E. 1994, AJ, 108, 2348
12. Ashman, K. M., & Zepf, S. E. 1992, ApJ, 384, 50
13. Ashman, K. M., & Zepf, S. E. 1998, Globular Clusters (New York: Cambridge University Press)
14. Baade, W. 1944, ApJ, 100, 147
15. Baade, W. 1958, in Stellar Populations, Proceedings of a Conference at Vatican Observatory, ed. D. J. K. O'Connell (New York: Interscience)
16. Babul, A., & Rees, M. J. 1992, MNRAS, 255, 346
17. Bahcall. N. A., Fan, X., & Cen, R. 1997, ApJ, 485, L53
18. Barbuy, B., Bica, E., & Ortolani, S. 1998, AAp, 333, 117
19. Barnes, J. E. 1988, ApJ, 331, 699
20. Barth, A. J., Ho, L. C., Filippenko, A. V, & Sargent, W. L. W. 1995, AJ, 110, 1009
21. Battistini, P. et al. 1987, AApS, 67, 447
22. Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 1998, ApJ, 498, 504
23. Baum, W. A. 1952, AJ, 57, 222
24. Baum, W. A. 1955, PASP, 67, 328
25. Baum, W. A. et al. 1995, AJ, 110, 2537
26. Baum, W. A. et al. 1997, AJ, 113, 1483
27. Baumgardt, H. 1998, AAp, 330, 480
28. Beck, S. C., Turner, J. L., Ho, P. T. P., Lacy, J. H., & Kelly, D. M. 1996, ApJ, 457, 610
29. Bender, R., Ziegler, B., & Bruzual, G. 1996, ApJ, 463, L51
30. Bernardi, M. et al. 1998, ApJ, 508, L143
31. Bershady, M. A., Hereld, M., Kron, R. G., Koo, D. C., Munn, J. A., & Majewski, S. R. 1994, AJ, 108, 870
32. Bertin, E., & Arnouts, S. 1996, AApS, 117, 393
33. Binggeli, B., & Popescu, C. C. 1995, AAp, 298, 63
34. Binggeli, B., Popescu, C. C., & Tammann, G. A. 1993, AApS, 98, 275
35. Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
36. Blakeslee, J. P. 1997, ApJ, 481, L59
37. Blakeslee, J. P., & Tonry, J. L. 1996, ApJ, 465, L19
38. Blakeslee, J. P., Tonry, J. L., & Metzger, M. R. 1997, AJ, 114, 482
39. Bothun G. D. 1986, AJ, 91, 507

40. Brewer, J. P., Richer, H. B., & Crabtree, D. R. 1995, AJ, 109, 2480
41. Bridges, T. J., Carter, D., Harris, W. E., & Pritchet, C. J. 1996, MNRAS, 281, 1290
42. Bridges, T. J., & Hanes, D. A. 1992, AJ, 103, 800 AJ, 101, 469
43. Bridges, T. J., Hanes, D. A., & Harris, W. E. 1991, AJ, 101, 469
44. Bridges, T. J. et al. 1997, MNRAS, 284, 376
45. Bridges, T. J., & Irwin, J. A. 1998, MNRAS, 300, 967
46. Brodie, J. P., & Huchra, J. P. 1991, ApJ, 379, 157
47. Brodie, J. P., Schroder, L. L., Huchra, J. P., Phillips, A. C., Kissler-Patig, M., & Forbes, D. A. 1998, AJ, 116, 691
48. Bruzual, G., & Charlot, S. 1993, ApJ, 405, 538
49. Buonanno, R., Corsi, C. E., Zinn, R., Fusi Pecci, F., Hardy, E., & Suntzeff, N. B. 1998, ApJ, 501, L33
50. Buonanno, R., & Iannicola, G. 1989, PASP, 101, 294
51. Burkert, A., & Smith, G. H. 1997, ApJ, 474, L15
52. Caldwell, N., & Bothun, G. D. 1987, AJ, 94, 1126
53. Calzetti, D. et al. 1997, AJ, 114, 1834
54. Capriotti, E. R., & Hawley, S. L. 1996, ApJ, 464, 765
55. Carlberg, R. G., & Pudritz, R. E. 1990, MNRAS, 247, 353
56. Carlberg, R. G. et al. 1996, ApJ, 462, 32
57. Carlson, M. N. et al. 1998, AJ, 115, 1778
58. Carney, B. W., & Seitzer, P. 1986, AJ, 92, 23
59. Carney, B. W., Storm, J., & Jones, R. V. 1992, ApJ, 386, 663
60. Carretta, E., & Gratton, R. G. 1997, AApS, 121, 95
61. Carretta, E., Gratton, R. G., Clementini, G., & Fusi Pecci, F. 1999, preprint (astro-ph/9902086)
62. Carroll, S. M., & Press, W. H. 1992, ARAA, 30, 499
63. Cassisi, S., Castellani, V., Degl'Innocenti, S., Salaris, M., & Weiss, A. 1999, AApS, 134, 103
64. Catelan, M., & de Freitas Pacheco, J. A. 1995, AAp, 297, 345
65. Cesarsky, D. A., Lequeux, J., Laustsen, S., Schuster, H.-E., & West, R. M. 1977, AAp, 61, L31
66. Chaboyer, B., Demarque, P., Kernan, P. J., & Krauss, L. M. 1998, ApJ, 494, 96
67. Chaboyer, B., Demarque, P., & Sarajedini, A. 1996, ApJ, 459, 558
68. Chiosi, C., Bertelli, G., & Bressan, A. 1992, ARAA, 30, 235
69. Christian, C. A., & Heasley, J. N. 1988, AJ, 95, 1422
70. Christian, C. A., & Heasley, J. N. 1991, AJ, 101, 848
71. Ciardullo, R., Jacoby, G. H., Feldmeier, J. J., & Bartlett, R. E. 1998, ApJ, 492, 62
72. Ciardullo, R., Jacoby, G. H., & Ford, H. C. 1989, ApJ, 344, 715
73. Ciardullo, R., Jacoby, G. H., & Tonry, J. L. 1993, ApJ, 419, 479
74. Cohen, J. G. 1992, ApJ, 400, 528
75. Cohen, J. G., Blakeslee, J. P., & Ryzhov, A. 1998, ApJ, 496, 808
76. Cohen, J. G., & Ryzhov, A. 1997, ApJ, 486, 230
77. Colless, M., & Dunn, A. M. 1996, ApJ, 458, 435
78. Conti, P. S., & Vacca, W. D. 1994, ApJ, 423, L97
79. Cool, A. M., Piotto, G., & King, I. R. 1996, ApJ, 468, 655
80. Côté, P. 1999, AJ, in press

81. Côté, P., Marzke, R. O., & West, M. J. 1998, ApJ, 501, 554
82. Cowan, J. J., McWilliam, A., Sneden, C., & Burris, D. L. 1997, ApJ, 480, 246
83. Couture, J., Harris, W. E., & Allwright, J. W. B. 1990, ApJS, 73, 671
84. Couture, J., Racine, R., Harris, W. E., & Holland, S. 1995, AJ, 109, 2050
85. Crampton, D., Cowley, A. P., Schade, D., & Chayer, P. 1985, ApJ, 288, 494
86. Cudworth, K. M. 1985, AJ, 90, 65
87. Cudworth, K. M., & Hanson, R. B. 1993, AJ, 105, 168
88. Da Costa, G. S. 1984, ApJ, 285, 483
89. Da Costa, G. S., & Armandroff, T. E. 1995, AJ, 1090, 2533
90. Dekel, A., & Silk, J. 1986, ApJ, 303, 39
91. De Marchi, G., Clampin, M., Greggio, L., Leitherer, C., Nota, A., & Tosi, M. 1997, ApJ, 479, L27
92. de Vaucouleurs, G. 1970, ApJ, 159, 435
93. de Vaucouleurs, G. 1977, Nature, 266, 126
94. Dinescu, D. I., Girard, T. M., & van Altena, W. F. 1999, preprint
95. Dorman, B., Lee, Y.-W., & VandenBerg, D. A. 1991, ApJ, 366, 115
96. Durrell, P. R., Harris, W. E., Geisler, D., & Pudritz, R. E. 1996a, AJ, 112, 972
97. Durrell, P. R., McLaughlin, D. E., Harris, W. E., & Hanes, D. A. 1996b, ApJ, 463, 543
98. Eckart, A. et al. 1990, ApJ, 363, 451
99. Eggen, O. J., Lynden-Bell, D., & Sandage, A. 1962, ApJ, 136, 748
100. Ellis, R. S. et al. 1997, ApJ, 483, 502
101. Elmegreen, B. G., & Efremov, Y. N. 1997, ApJ, 480, 235
102. Elmegreen, B. G., Efremov, Y. N., Pudritz, R. E., & Zinnecker, H. 1999, in Protostars and Planets IV, ed. A. P. Boss, S. S. Russell, & V. Mannings (Tucson: Univ.Arizona Press), in press
103. Elmegreen, B. G., & Falgarone, E. 1996, ApJ, 471, 816
104. Elson, R. A. W., Grillmair, C. J., Forbes, D. A., Rabban, M., Williger, G. M., & Brodie, J. P. 1998, MNRAS, 295, 240
105. Fabian, A. C., Nulsen, P. E. J., & Canizares, C. R. 1984, MNRAS, 201, 933
106. Fall, S. M., & Rees, M. J. 1977, MNRAS, 181, 37P
107. Fall, S. M., & Rees, M. J. 1985, ApJ, 298, 18
108. Feast, M. 1998, MNRAS, 293, L27
109. Feldmeier, J. J., Ciardullo, R., & Jacoby, G. H. 1997, ApJ, 479, 231
110. Ferguson, H. C., & Sandage, A. 1989, ApJ, 346, L53
111. Ferguson, H. C., Tanvir, N. R., & von Hippel, T. 1998, Nature, 391, 461
112. Fernley, J. et al. 1998a, AAp, 330, 515
113. Fernley, J. et al. 1998b, MNRAS, 293, L61
114. Ferrarese, L. et al. 1996, ApJ, 464, 568
115. Ferraro, F. R. et al. 1997, AAp, 320, 757
116. Field, G. B., & Saslaw, W. C. 1965, ApJ, 142, 568
117. Fleming, D. E. B., Harris, W. E., Pritchet, C. J., & Hanes, D. A. 1995, AJ, 109, 1044
118. Forbes, D. A. 1996a, AJ, 112, 954
119. Forbes, D. A. 1996b, AJ, 112, 1409
120. Forbes, D. A., Brodie, J. P., & Huchra, J. 1996a, AJ, 112, 2448
121. Forbes, D. A., Brodie, J. P., & Grillmair, C. J. 1997, AJ, 113, 1652
122. Forbes, D. A., Franx, M., Illingworth, G. D., & Carollo, C. M. 1996b, ApJ, 467, 126

123. Forbes, D. A., Grillmair, C. J., Williger, G. M., Elson, R. A. W., & Brodie, J. P. 1998, MNRAS, 293, 325
124. Ford, H. C., Hui, X., Ciardullo, R., Jacoby, G. H., & Freeman, K. C. 1996, ApJ, 458, 455
125. Forte, J. C., Strom, S. E., & Strom, K. M. 1981, ApJ, 245, L9
126. Freedman, W. L., & Madore, B. F. 1990, ApJ, 365, 186
127. Frenk, C. S., & White, S. D. M. 1980, MNRAS, 193, 295
128. Fritze-von Alvensleben, U. 1998, AAp, 336, 83
129. Fritze-von Alvensleben, U. 1999, AAp, 342, L25
130. Frogel, J. A., Cohen, J. G., & Persson, S. E. 1983, ApJ, 275, 773
131. Fruchter, A. S., & Hook, R. N. 1999, PASP, 111, in press (astro-ph/9808087)
132. Fusi Pecci, F., Bellazzini, M., Cacciari, C., & Ferraro, F. R. 1995, AJ, 110, 1664
133. Fusi Pecci, F. et al. 1996, AJ, 112, 1461
134. Geisler, D., & Forte, J. C. 1990, ApJ, 350, L5
135. Geisler, D., Lee, M. G., & Kim, E. 1996, AJ, 111, 1529
136. Giavalisco, M., Steidel, C. C., & Macchetto, F. D. 1996, ApJ, 474, 223
137. Gieren, W. P., Fouqué, P., & Gómez, M. 1998, ApJ, 496, 17
138. Gilliland, R. L. 1992, in Astronomical CCD Observing and Reduction Techniques, ASP Conf.Series 23, ed. S. B. Howell (San Francisco: ASP), 68
139. Girardi, M., Biviano, A., Giuricin, G., Mardirossian, F., & Mezzetti, M. 1993, ApJ, 404, 38
140. Glazebrook, K., Abraham, R., Santiago, B., Ellis, R., & Griffiths, R. 1998, MNRAS, 297, 885
141. Gnedin, O. Y. 1997, ApJ, 487, 663
142. Gnedin, O. Y., & Ostriker, J. P. 1997, ApJ, 474, 223
143. Gonzalez Delgado, R. M., Leitherer, C., Heckman, T., & Cervino, M. 1997, ApJ, 483, 705
144. Gorjian, V. 1996, AJ, 112, 1886
145. Gould, A. 1994, ApJ, 426, 542
146. Gould, A., & Popowski, P. 1998, ApJ, 508, 844
147. Graham, J. A. 1977, PASP, 89, 425
148. Graham, J. A. et al. 1997, ApJ, 477, 535
149. Gratton, R. G. 1998, MNRAS, 296, 739
150. Gratton, R. G., Fusi Pecci, F., Carretta, E., Clementini, G., Corsi, C. E., & Lattanzi, M. 1997, ApJ, 491, 749
151. Grillmair, C. J. 1998, in Galactic Halos: A UC Santa Cruz Workshop, ASP Conf. Proc. 136, ed. D. Zaritsky (San Francisco: ASP), 45
152. Grillmair, C. J. et al. 1994, ApJ, 422, L9
153. Grillmair, C. J., Forbes, D. A., Brodie, J. P., & Elson, R. A. W. 1999, AJ, 117, 167
154. Grundahl, F., VandenBerg, D. A., & Andersen, M. I. 1998, ApJ, 500, L179
155. Gullixson, C. A. 1992, in Astronomical CCD Observing and Reduction Techniques, ASP Conf.Series 23, ed. S. B. Howell (San Francisco: ASP), 130
156. Hamuy, M., Phillips, M. M., Suntzeff,N. B., Schommer, R. A., Maza, J., & Avilés, R. 1996, AJ, 112, 2398
157. Hanes, D. A. 1977, MNRAS, 180, 309
158. Hanes, D. A., & Whittaker, D. G. 1987, AJ, 94, 906
159. Hanson, R. B. 1979, MNRAS, 186, 875

160. Harris, G. L. H., Geisler, D., Harris, H. C., & Hesser, J. E. 1992, AJ, 104, 613
161. Harris, G. L. H., Harris, W. E., & Poole, G. B. 1999, AJ, 117, 855
162. Harris, G. L. H., Poole, G. B., & Harris, W. E. 1998, AJ, 116, 2866
163. Harris, H. C. 1988, in Globular Cluster Systems in Galaxies, IAU Symposium 126, ed. J. E.Grindlay & A. G. D. Philip (Dordrecht: Kluwer), 205
164. Harris, W. E. 1976, AJ, 81, 1095
165. Harris, W. E. 1981, ApJ, 251, 497
166. Harris, W. E. 1988a, in Globular Cluster Systems in Galaxies, IAU Symposium 126, ed. J. E.Grindlay & A. G. D. Philip (Dordrecht: Kluwer), 237
167. Harris, W. E. 1988b, in The Extragalactic Distance Scale, ASP Conf.Series 4, ed. S. van den Bergh & C. J. Pritchet (San Francisco: ASP), 231
168. Harris, W. E. 1990a, PASP, 102, 949
169. Harris, W. E. 1990b, PASP, 102, 966
170. Harris, W. E. 1991, ARAA, 29, 543
171. Harris, W. E. 1993, in The Globular Cluster – Galaxy Connection, ASP Conf. Series 48, ed. G. H. Smith & J. P. Brodie (San Francisco: ASP), 472
172. Harris, W. E. 1995, in Stellar Populations, IAU Symposium 164, ed. P. C. van der Kruit & G. Gilmore (Dorcrecht: Kluwer), 85
173. Harris, W. E. 1996a, AJ, 112, 1487
174. Harris, W. E. 1996b, in Formation of the Galactic Halo – Inside and Out, ASP Conf. Proc. 92, ed. H. Morrison & A. Sarajedini (San Francisco: ASP), 231
175. Harris, W. E. 1998, in Galactic Halos: A UC Santa Cruz Workshop, ASP Conf. Proc. 136, ed. D. Zaritsky (San Francisco: ASP), 33
176. Harris, W. E. 1999, in Globular Clusters, Tenth Canary Islands Winter School (Cambridge University Press), in press
177. Harris, W. E., Allwright, J. W. B., Pritchet, C. J., & van den Bergh, S. 1991, ApJS, 76, 115
178. Harris, W. E., & Canterna, R. 1979, ApJ, 231, L19
179. Harris, W. E. et al. 1997a, AJ, 113, 688
180. Harris, W. E. et al. 1997b, AJ, 114, 1030
181. Harris, W. E., Durrell, P. R., Petitpas, G. R., Webb, T. M., & Woodworth, S. C. 1997c, AJ, 114, 1043
182. Harris, W. E., Durrell, P. R., Pierce, M. J., & Secker, J. 1998b, Nature, 395, 45
183. Harris, W.E., FitzGerald, M. P., & Reed, B. C. 1981, PASP, 93, 507
184. Harris, W. E., Harris, G. L. H., & McLaughlin, D. E. 1998a, AJ, 115, 1801
185. Harris, W. E., Pritchet, C. J., & McClure, R. D. 1995, ApJ, 441, 120
186. Harris, W. E., & Pudritz, R. E. 1994, ApJ, 429, 177
187. Harris, W. E., & Racine, R. 1979, ARAA, 17, 241
188. Harris, W. E., & Smith, M. G. 1976, ApJ, 207, 1036
189. Harris, W. E., & van den Bergh, S. 1981, AJ, 86, 1627
190. Hartwick, F. D. A. 1976, ApJ, 209, 418
191. Hartwick, F. D. A. 1996, in Formation of the Galactic Halo – Inside and Out, ASP Conf. Proc. 92, ed. H. Morrison & A. Sarajedini (San Francisco: ASP), 444
192. Hartwick, F. D. A., & Sargent, W. L. W. 1978, ApJ, 221, 512
193. Heasley, J. N., Christian, C. A., Friel, E. D., & Janes, K. A. 1988, AJ, 96, 1312
194. Hesser, J. E., Harris, W. E., VandenBerg, D. A., Allwright, J. W. B., Shott, P., & Stetson, P. B. 1987, PASP, 99, 739

195. Hilker, M. 1998, PhD thesis, University of Bonn
196. Hodge, P. W. 1974, PASP, 86, 289
197. Holland, S. 1998, AJ, 115, 1916
198. Holland, S., Fahlman, G. G., & Richer, H. B. 1997, AJ, 114, 1488
199. Holtzman, J. A. et al. 1992, AJ, 103, 691
200. Holtzman, J. A. et al. 1996, AJ, 112, 416
201. Hook, R. N., & Fruchter, A. S. 1997, in Astronomical Data Analysis Software and Systems VI, ASP Conf.Series 125, ed. G. Hunt & H. E. Payne (San Francisco: ASP), 147
202. Howell, S. B. 1992, in Astronomical CCD Observing and Reduction Techniques, ASP Conf.Series 23, ed. S. B. Howell (San Francisco: ASP), 105
203. Hubble, E. 1932, ApJ, 76, 44
204. Huchra, J. P. 1988, in The Extragalactic Distance Scale, ASP Conf.Series 4, ed. S. van den Bergh & C. J. Pritchet (San Francisco: ASP), 257
205. Huchra, J. P., & Brodie, J. P. 1987, AJ, 93, 779
206. Huchra, J. P., Brodie, J. P., & Kent, S. M. 1991, ApJ, 370, 495
207. Hui, X., Ford, H. C., Freeman, K. C., & Dopita, M. A. 1995, ApJ, 449, 592
208. Hunter, D. A., O'Connell, R. W., & Gallagher, J. S. 1994a, AJ, 108, 84
209. Hunter, D. A., van Woerden, H., & Gallagher, J. S. 1994b, ApJS, 91, 79
210. Infante, L. 1987, AAp, 183, 177
211. Infante, L., & Pritchet, C. J. 1992, ApJS, 82, 237
212. Irwin, M. J., Demers, S., & Kunkel, W. E. 1995, ApJ, 453, L21
213. Israel, F. P. 1997, AAp, 328, 471
214. Jablonka, P., Alloin, D., & Bica, E. 1992, AAp, 260, 97
215. Jablonka, P., Bica, E., Pelat, D., & Alloing, D. 1996, AAp, 307, 385
216. Jacoby, G. H., Ciardullo, R., & Ford, H. C. 1990, ApJ, 356, 332
217. Jacoby, G. H., Ciardullo, R., & Harris, W. E. 1996, ApJ, 462, 1
218. Jacoby, G. H. et al. 1992, PASP, 104, 599
219. Jaffe, W. 1990, AAp, 240, 254
220. Jarvis, J. F., & Tyson, J. A. 1981, AJ, 86, 476
221. Jerjen, H., & Tammann, G. A. 1993, AAp, 276, 1
222. Johnson, J. A., & Bolte, M. 1998, AJ, 115, 693
223. Johnson, J. A., Bolte, M., Stetson, P. B., & Hesser, J. E. 1999, ApJ, submitted
224. Johnston, K. V. 1998, in Galactic Halos: A UC Santa Cruz Workshop, ASP Conf. Proc. 136, ed. D. Zaritsky (San Francisco: ASP), 365
225. Jones, B. F., Klemola, A. R., & Lin, D. N. C. 1994, AJ, 107, 1333
226. Kavelaars, J. J. 1998, PASP, 110, 758
227. Kavelaars, J. J. 1999, preprint (astro-ph/9806094)
228. Kavelaars, J. J., & Gladman, B. 1998, in Proc. 5th CFHT Users Meeting, ed. P. Martin & S. Rucinski (Kamuela: CFHT), 155
229. Kavelaars, J. J., & Hanes, D. A. 1997, MNRAS, 285, L31
230. Kavelaars, J. J., Harris, W. E., Pritchet, C. J., Hanes, D. A., & Hesser, J. E. 1999, in preparation
231. King, I. R., Anderson, J., Cool, A., & Piotto, G. 1998, ApJ, 492, L37
232. Kinman, T. D. 1959a, MNRAS, 119, 538
233. Kinman, T. D. 1959b, MNRAS, 119, 559
234. Kinman, T. D. et al. 1991, PASP, 103, 1279
235. Kissler, M., Richtler, T., Held, E. V., Grebel, E. K., Wagner, S. J., & Capaccioli, M. 1994, AAp, 287, 463

236. Kissler-Patig, M. 1998, in Galactic Halos, ASP Conference Series 136, ed. D. Zaritsky (San Francisco: ASP), 63
237. Kissler-Patig, M., Ashman, K. M., Zepf, S. E., & Freeman, K. C. 1999, AJ, in press (astro-ph/9903324)
238. Kissler-Patig, M., Forbes, D. A., & Minniti, D. 1998, MNRAS, 298, 1123
239. Kissler-Patig, M., & Gebhardt, K. 1998, AJ, 116 2237
240. Kissler-Patig, M., Grillmair, C. J., Meylan, G., Brodie, J. P., Minniti, D., & Goudfrooij, P. 1999, AJ, 117, 1206
241. Kissler-Patig, M., Kohle, S., Hilker, M., Richtler, T., Infante, L., & Quintana, H. 1997a, AAp, 319, 470
242. Kissler-Patig, M., Richtler, T., Storm, J., & Della Valle, M. 1997b, AAp, 327, 503
243. Kobulnicky, H. A., Dickey, J. M., Sargent, A. I., Hogg, D. E., & Conti, P. S. 1995, AJ, 110, 117
244. Kobulnicky, H. A., & Skillman, E. D. 1996, ApJ, 471, 211
245. Kohle, S., Kissler-Patig, M., Hilker, M., Richtler, T., Infante, L., & Quintana, H. 1996, AAp, 309, L39
246. Kron, G. E., & Mayall, N. U. 1960, AJ, 65, 581
247. Kron, R. 1980, ApJS, 43, 305
248. Kroupa, P., & Bastian, U. 1997a, New Astron, 2, 77
249. Kroupa, P., & Bastian, U. 1997b, in Proc.ESA Symposium "Hipparcos – Venice '97", ESA SP-402, 615
250. Kundu, A., & Whitmore, B. C. 1998, AJ, 116, 2841
251. Kundu, A., Whitmore, B. C., Sparks, W. B., Macchetto, F. D., Zepf, S. E., & Ashman, K. M. 1999, ApJ, 513, 733
252. Kunkel, W. E., & Demers, S. 1975, RGO Bulletin No. 182
253. Kunkel, W. E., & Demers, S. 1977, ApJ, 214, 21
254. Kwan, J. 1979, ApJ, 229, 567
255. Larson, R. B. 1988, in Globular Cluster Systems in Galaxies, IAU Symposium No. 126, ed. J. E. Grindlay & A. G. D. Philip (Dordrecht: Kluwer), 311
256. Larson, R. B. 1990a, in Physical Processes in Fragmentation and Star Formation, ed. R. Capuzzo-Dolcetta, C. Chiosi, & A. DiFazio (Dordrecht: Kluwer), 389
257. Larson, R. B. 1990b, PASP, 653, 704
258. Larson, R. B. 1993, in The Globular Cluster - Galaxy Connection, ASP Conf.Ser. 48, ed. G. H. Smith & J. P. Brodie (San Francisco: ASP), 675
259. Larson, R. B. 1996, in Formation of the Galactic Halo ... Inside and Out, ASP Conf.Ser., 92, ed. H. Morrison & A. Sarajedini (San Francisco: ASP), 241
260. Lauberts, A. 1976, AAp, 52, 309
261. Lauer, T. R. 1999, PASP, 111, 227
262. Lauer, T. R., Tonry, J. R., Postman, M., Ajhar, E. A., & Holtzman, J. A. 1998, ApJ, 499, 577
263. Layden, A. C., Hanson, R. B., Hawley, S. L., Klemola, A. R., & Hanley, C. J. 1996, AJ, 112, 2110
264. Lazareff, B., Castets, A., Kim, D. W., & Jura, M. 1989, ApJ, 336, L13
265. Lee, M. G., Freedman, W. L., & Madore, B. F. 1993a, ApJ, 417, 553
266. Lee, M. G., Freedman, W. L., Mateo, M., Thompson, I., Roth, M., & Ruiz, M. T. 1993b, AJ, 106, 1420
267. Lee, M. G., Kim, E., & Geisler, D. 1997, AJ, 114, 1824

268. Lee, M. G., Kim, E., & Geisler, D. 1998, AJ, 115, 947
269. Lee, Y.-W., Demarque, P., & Zinn, R. 1994, ApJ, 423, 248
270. Leitherer, C., Vacca, W. D., Conti, P. S., Filippenko, A. V., Robert, C., & Sargent, W. L. W. 1996, ApJ, 465, 717
271. Lin, D. N. C., Jones, B. F., & Klemola, A. R. 1995, ApJ, 439, 652
272. Lineweaver, C. H., Barbosa, D., Blanchard, A., & Bartlett, J. G. 1998, AAp, 322, 365
273. Little, B., & Tremaine, S. 1987, ApJ, 320, 493
274. Liu, T., & Janes, K. A. 1990, ApJ, 360, 561
275. Lo, K. Y., Cheung, K. W., Masson, C. R., Phillips, T. G., Scott, S. L., & Woody, D. P. 1987, ApJ, 312, 574
276. Long, K., Ostriker, J. P., & Aguilar, L. 1992, ApJ, 388, 362
277. Loewenstein, M. & Mushotzky, R. F. 1996, ApJ, 466, 695
278. Lutz, D. 1991, AAp, 245, 31
279. Lutz, T. E., & Kelker, D. H. 1973, PASP, 85, 573
280. Lynden-Bell, D. 1976, MNRAS, 174, 695
281. Lynden-Bell, D., Cannon, R. D., & Godwin, P. J. 1983, MNRAS, 204, 87P
282. Lynden-Bell, D., & Lynden-Bell, R. M. 1995, MNRAS, 275, 429
283. Madau, P., Ferguson, H. C., Dickinson, M.E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388
284. Madore, B. F., & Arp, H. C. 1979, ApJ, 227, L103
285. Madore, B. F. et al. 1998, Nature, 395, 47
286. Majewski, S. R. 1993, ARAA, 31, 575
287. Majewski, S. R. 1994, ApJ, 431, L17
288. Majewski, S. R., Hawley, S. L., & Munn, J. A. 1996, in Formation of the Galactic Halo ... Inside and Out, ASP Conf. Series 92, edited by H. Morrison & A. Sarajedini (San Francisco: ASP), 119
289. Massey, P., & Jacoby, G. H. 1992, in Astronomical CCD Observing and Reduction Techniques, ASP Conf.Series 23, ed. S. B. Howell (San Francisco: ASP), 240
290. Maoz, E. 1990, ApJ, 359, 257
291. Marsakov, V. A., & Suchkov, A. A. 1976, Sov.Astron.Letters, 2, 148
292. Massey, P., & Hunter, D. A. 1998, ApJ, 493, 180
293. Mateo, M. 1996, in Formation of the Galactic Halo ... Inside and Out, edited by H. Morrison & A. Sarajedini (San Francisco: ASP), 434
294. Mateo, M., Fischer, P, & Krzeminski, W. 1995, AJ, 110, 2166
295. Mateo, M., Olszewski, E. W., Pryor, C., Welch, D. L., & Fischer, P. 1993, AJ, 105, 510
296. Mateo, M., Olszewski, E. W., Welch, D. L., Fischer, P., & Kunkel, W. 1991, AJ, 102, 914
297. Mathews, G. J., Kajino, T., & Orito, M. 1996, ApJ, 456, 98
298. Matthews, L. D., Gallagher, J. S., & Littleton, J. E. 1995, AJ, 110, 581
299. Mauersberger, R., Henkel, C., Wielibinski, R., Wiklind, T., & Reuter, H. P. 1996, AAp, 305, 421
300. Mayall, N. U. 1946, ApJ, 104, 290
301. Mayall, N. U., & Eggen, O. J. 1953, PASP, 65, 24
302. Mayya, Y. D., & Prabhu, T. P. 1996, AJ, 111, 1252
303. McKee, C.F., Zweibel, E. G., Goodman, A. A., and Heiles, C. 1993, in Protostars and Planet III, ed. M. Matthews & E. Levy (Tucson: University of Arizona Press), 327.

304. McLaughlin, D. E. 1994, PASP, 106, 47
305. McLaughlin, D. E. 1999, AJ, in press
306. McLaughlin, D. E., Harris, W. E., & Hanes, D. A. 1994, ApJ, 422, 486
307. McLaughlin, D. E., & Pudritz, R. E. 1996, ApJ, 457, 578
308. McMillan, R., Ciardullo, R., & Jacoby, G. H. 1993, ApJ, 416, 62
309. McNamara, B. R., & O'Connell, R. W. 1992, ApJ, 393, 579
310. Meurer, G. R. 1995, Nature, 375, 742
311. Meurer, G. R., Freeman, K. C., Dopita, M. A., & Cacciari, C. 1992, AJ, 103, 60
312. Meurer, G. R., Heckman, T. M., Leitherer, C., Kinney, A., Robert, C., & Garnett, D. R. 1995, AJ, 110, 2665
313. Mighell, K. J., & Rich, R. M. 1995, AJ, 110, 1649
314. Mighell, K. J., & Rich, R. M. 1996, AJ, 111, 777
315. Miller, B. W., Lotz, J. M., Ferguson, H. C., Stiavelli, M., & Whitmore, B. C. 1998, ApJ, 508, L133
316. Miller, B. W., Whitmore, B. C., Schweizer, F., & Fall, S. M. 1997, AJ, 114, 2381
317. Minniti, D. 1995, AJ, 109, 1663
318. Minniti, D., Alonso, M. V., Goudfrooij, P., Jablonka, P., & Meylan, G. 1996, ApJ, 467, 221
319. Minniti, D., Kissler-Patig, M, Goudfrooj, P., & Meylan, G. 1998, AJ, 115, 121
320. Morgan, W. W. 1956, PASP, 68, 509
321. Morris, P. W., & Shanks, T. 1998, MNRAS, 298, 451
322. Mould, J. R., Oke, J. B., De Zeeuw, P. T., & Nemec, J. M. 1990, AJ, 99, 1823
323. Mould, J. R., Oke, J. B., & Nemec, J. M. 1987, AJ, 93, 53
324. Mould, J. et al. 1995, ApJ, 449, 413
325. Murali, C., & Weinberg, M. D. 1997a, MNRAS, 288, 749
326. Murali, C., & Weinberg, M. D. 1997b, MNRAS, 288, 767
327. Murali, C., & Weinberg, M. D. 1997c, MNRAS, 291, 717
328. Mushotzky, R. F., & Loewenstein, M. 1997, ApJ, 481, L63
329. Muzzio, J. C. 1986, ApJ, 306, 44
330. Muzzio, J. C. 1988, in Globular Cluster Systems in Galaxies, IAU Symposium 126, edited by J. Grindlay & A. G. D. Philip (Dordrecht: Reidel), 543
331. Navarro, J. F., Frenk, C., & White, S. D. M. 1996, ApJ, 462, 563
332. Neilsen, E. H. Jr, Tsvetanov, Z. I., & Ford, H. C. 1997 ApJ, 483, 745
333. Nemec, J. M., Wehlau, A., & de Oliviera, C. M. 1988, AJ, 96, 528
334. Norris, J. E. 1993, in The Globular Cluster – Galaxy Connection, ASP Conf. Series 48, edited by G. Smith & J. P. Brodie (San Francisco: ASP), 259
335. Norris, J. E., & Hawkins, M. R. S. 1991, ApJ, 380, 104
336. O'Connell, R. W., Gallagher, J. S., & Hunter, D. A. 1994, ApJ, 433, 65
337. O'Connell, R. W., Gallagher, J. S., Hunter, D. A., & Colley, W. N. 1995, ApJ, 446, L1
338. Odenkirchen, M., Brosche, P., Geffert, M., & Tucholke, H.-J. 1997, New Astron, 2, 477
339. Okazaki, T., & Tosa, M. 1995, MNRAS, 274, 48
340. Olsen, K. A. G., et al. 1998, MNRAS, 300, 665
341. Oort, J. H. 1977, ApJ, 218, L97
342. Ostriker, J. P., & Gnedin, O. Y. 1997, ApJ, 487, 667
343. Ostrov, P. G., Forte, J. C., & Geisler, D. 1998, AJ, 116, 2854

344. Pahre, M. A., & Mould, J. R. 1994, ApJ, 433, 567
345. Pascarelle, S. M., Windhorst, R. A., Keel, W. C., & Odewahn, S. C. 1996, Nature, 383, 45
346. Peebles, P. J. E. 1993, Principles of Physical Cosmology (Princeton: Princeton University Press)
347. Peebles, P. J. E., & Dicke, R. H. 1968, ApJ, 154, 891
348. Perelmuter, J.-M., & Racine, R. 1995, AJ, 109, 1055
349. Perlmutter, S. et al. 1997, ApJ, 483, 565
350. Pierce, M. J., Welch, D. L., McClure, R. D., van den Bergh, S., Racine, R., & Stetson, P. B. 1994, Nature, 371, 385
351. Pont, F., Mayor, M., Turon, C., & VandenBerg, D. A. 1998, AAp, 329, 87
352. Pritchet, C. J., & Kline, M. I. 1981, AJ, 86, 1859
353. Pritchet, C. J., & van den Bergh, S. 1987, ApJ, 316, 517
354. Pritchet, C. J., & van den Bergh, S. 1994, AJ, 107, 1730
355. Puzia, T. H., Kissler-Patig, M., & Brodie, J. P. 1999, poster paper at Globular Clusters, Tenth Canary Islands Winter School
356. Queloz, D., Dubath, P, & Pasquini, L. 1995, AAp, 300, 31
357. Racine, R. 1968, JRASC, 62, 367
358. Racine, R. 1980, in Star Clusters, IAU Symposium 85, ed. J. E. Hesser (Dordrecht: Reidel), 369
359. Racine, R. 1991, AJ, 101, 865
360. Racine, R., & Harris, W. E. 1975, ApJ, 196, 413
361. Racine, R., & Harris, W. E. 1989, AJ, 98, 1609
362. Racine, R., & Harris, W. E. 1992, AJ, 104, 1068
363. Racine, R., Oke, J. B., & Searle, L. 1978, ApJ, 223, 82
364. Reed, L., Harris, G. L. H., & Harris, W. E. 1992, AJ, 103, 824
365. Reed, L., Harris, G. L. H., & Harris, W. E. 1994, AJ, 107, 555
366. Rees, R. F. 1996, in Formation of the Galactic Halo ... Inside and Out, ASP Conf. Series 92, ed. H. Morrison & A. Sarajedini (San Francisco: ASP), 289
367. Rees, R. F., & Cudworth, K. M. 1991, AJ, 102, 152
368. Reid, I. N. 1997, AJ, 114, 161
369. Reid, I. N., & Freedman, W. 1994, MNRAS, 267, 821
370. Renzini, A., & Fusi Pecci, F. 1988, ARAA, 26, 199
371. Renzini, A. et al. 1996, ApJ, 465, L23
372. Rich, R. M., Mighell, K. J., Freedman, W. L., & Neill, J. D. 1996, AJ, 111, 768
373. Richer, H. B. et al. 1995, ApJ, 451, L17
374. Richer, H. B. et al. 1997, ApJ, 484, 741
375. Richtler, T. 1992, Habilitationsschrift zur Erlangung, Sternwarte der Universitat Bonn
376. Richtler, T., Grebel, E. K., Domgorgen, H., Hilker, M., & Kissler, M. 1992, AAp, 264, 25
377. Riess, A. G. et al. 1998, AJ, 116, 1009
378. Rodgers, A. W., & Paltoglou, G. 1984, ApJ, 283, L5
379. Roman, N. G. 1952, ApJ, 116, 122
380. Rutledge, G. A., Hesser, J. E., & Stetson, P. B. 1997, PASP, 109, 907
381. Ryden, B. S., & Terndrup, D. M. 1994, ApJ, 425, 43
382. Saha, A., Sandage, A., Labhardt, L., Tammann, G. A., Macchetto, F. D., & Panagia, N. 1996a, ApJ, 466, 55

383. Saha, A., Sandage, A., Labhardt, L., Tammann, G. A., Macchetto, F. D., & Panagia, N. 1996b, ApJS, 107, 693
384. Sakai, S., Madore, B. F., Freedman, W. L., Lauer, T. R., Ajhar, E. A., & Baum, W. A. 1997, ApJ, 478, 49
385. Salaris, M., Degl'Innocenti, S., & Weiss, A. 1997, ApJ, 479, 665
386. Sandage, A. 1968, ApJ, 152, L149
387. Sandage, A. 1970, ApJ, 162, 841
388. Sandquist, E. L., Bolte, M., Stetson, P. B., & Hesser, J. E. 1996, ApJ, 470, 910
389. Sargent, W. L. W., Kowal, S. T., Hartwick, F. D. A., & van den Bergh, S. 1977, AJ, 82, 947
390. Satypal, S. et al. 1997, ApJ, 483, 148
391. Schechter, P. L., Mateo, M., & Saha, A. 1993, PASP, 105, 1342
392. Schreier, E. J., Capetti, A., Maccheto, F. , Sparks, W. B., & Ford, H. J. 1996, ApJ, 459, 535
393. Schweizer, F. 1987, in Nearly Normal Galaxies, Eighth Santa Cruz Summer Workshop in Astronomy and Astrophysics, ed. S. M. Faber (New York: Springer), 18
394. Schweizer, F. 1996, AJ, 111, 109
395. Schweizer, F., Miller, B. W., Whitmore, B. C., & Fall, S. M. 1996, AJ, 112, 1839
396. Scoville, N. Z. 1998, ApJ, 492, L107
397. Scoville, N. Z., Soifer, B T., Neugebauer, G., Young, J. S., Matthews, K., & Yerka, J. 1985, ApJ, 289, 129
398. Searle, L. 1977, in The Evolution of Galaxies and Stellar Populations, edited by B. M. Tinsley & R. Larson (New Haven: Yale University Observatory), 219
399. Searle, L., & Zinn, R. 1978, ApJ, 225, 357
400. Secker, J. 1992, AJ, 104, 1472
401. Secker, J., Geisler, D., McLaughlin, D. E., & Harris, W. E. 1995, AJ, 109, 1019
402. Secker, J., & Harris, W. E. 1993, AJ, 105, 1358
403. Secker, J., & Harris, W. E. 1997, PASP, 109, 1364
404. Seyfert, C. K., & Nassau, J. J. 1945, ApJ, 102, 377
405. Shapley, H. 1918, ApJ, 48, 154
406. Sharov, A. S. 1976, Sov.Astron, 20, 397
407. Sharples, R. M. et al. 1998, AJ, 115, 2337
408. Shaya, E. J., Dowling, D. M., Currie, D. G., Faber, S. M., & Groth, E. J. 1994, AJ, 107, 1675
409. Simard, L., & Pritchet, C. J. 1994, AJ, 107, 503
410. Smecker-Hane, T. A., Stetson, P. B., Hesser, J. E., & Lehnert, M. D. 1994, AJ, 108, 507
411. Smith, E. O., Neill, J. D., Mighell, K. J., & Rich, R. M. 1996, AJ, 111, 1596
412. Sodemann, M., & Thomsen, B. 1996, AJ, 111, 208
413. Stanek, K. Z., & Garnavich, P. M. 1998, ApJ, 503, L131
414. Stanford, S. A., Eisenhardt, P. R., & Dickinson, M. 1998, ApJ, 492, 461
415. Stanford, S. A., Sargent, A. I., Sanders, D. B., & Scoville, N. Z. 1990, ApJ, 349, 492
416. Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., Pettini, M., & Kellogg, M. 1998, ApJ, 492, 428

417. Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17
418. Stetson, P. B. 1987, PASP, 99, 191
419. Stetson, P. B. 1990, PASP, 102, 932
420. Stetson, P. B. 1994, PASP, 106, 250
421. Stetson, P. B. 1998, private communication
422. Stetson, P. B. et al. 1999, AJ, 117, 247
423. Stetson, P. B., Davis, L. E., & Crabtree, D. R. 1990, in CCDs in Astronomy, ASP Conf Ser 8, edited by G. H. Jacoby (San Francisco: ASP), 289
424. Stetson, P. B., & Harris, W. E. 1988, AJ, 96, 909
425. Stewart, G. C., Fabian, A. C., Jones, C., & Forman, W. 1984, ApJ, 285, 1
426. Storm, J., Nordstrom, B., Carney, B. W., & Andersen, J. 1994a, AAp, 290, 443
427. Storm, J., Carney, B. W., & Latham, D. W. 1994b, AAp, 291, 121
428. Strom, S. E., Forte, J. C., Harris, W. E., Strom, K. M., Wells, D. C., & Smith, M. G. 1981, ApJ, 245, 416
429. Suntzeff, N. B., Mateo, M., Terndrup, D. M., Olszewski, E. W., Geisler, D., & Weller, W. 1993, ApJ, 418, 208
430. Surdin, V. G. 1979, Sov.Astron., 23, 648
431. Tacconi, L. J., & Young, J. S. 1985, ApJ, 290, 602
432. Theuns, T., & Warren, S. J. 1997, MNRAS, 284, L11
433. Tonry, J. L., Blakeslee, J. P., Ajhar, E. A., & Dressler, A. 1997, ApJ, 475, 399
434. Toomre, A. 1977, in The Evolution of Galaxies and Stellar Populations, edited by B. M. Tinsley & R. B. Larson (New Haven: Yale University Observatory), 401
435. Tremaine, S. D., Ostriker, J. P., & Spitzer, L. Jr 1975, ApJ, 196, 407
436. Turner, E. L. 1991, AJ, 101, 5
437. Turner, J. L., Beck, S. C., & Hunt, R. L. 1997, ApJ, 474, L11
438. Tyson, A. 1989, in CCDs in Astronomy, ASP Conf.Series 8, ed. G. H. Jacoby (San Francisco: ASP), 1
439. Tyson, J. A., & Jarvis, J. F. 1979, ApJ, 230, L153
440. Valdes, F., Tyson, J. A., & Jarvis, J. F. 1983, ApJ, 271, 431
441. van Altena, W. F., Lee, J. T., & Hoffleit, E. D. 1995, General Catalogue of Trigonometric Stellar Parallaxes, Fourth Edition (New Haven: Yale Univ. Observatory)
442. VandenBerg, D. A., Stetson, P. B., & Bolte, M. 1996, ARAA, 34, 461
443. van den Bergh, S. 1969, ApJS, 19, 145
444. van den Bergh, S. 1977, Vistas Astron, 21, 71
445. van den Bergh, S. 1982, PASP, 94, 459
446. van den Bergh, S. 1993a, AJ, 105, 971
447. van den Bergh, S. 1993b, ApJ, 411, 178
448. van den Bergh, S. 1995, ApJ, 446, 39
449. van den Bergh, S., Abraham, G. G., Ellis, R. S., Tanvir, N. R.,, Santiago, B. X., & Glazebrook, K. G. 1996, AJ, 112, 359
450. van den Bergh, S., Pritchet, C. J., & Grillmair, C. 1985, AJ, 90, 595
451. Vesperini, E. 1997, MNRAS, 287, 915
452. Vesperini, E. 1998, MNRAS, 299, 1019
453. Vesperini, E., & Heggie, D. C. 1997, MNRAS, 289, 898
454. Veteŝnik, M. 1962, Bull.Astron.Inst.Czech., 13, 180 and 218

455. Vogt, S. S., Mateo, M., Olszewski, E. W., & Keane, M. J. 1995, AJ, 109, 151
456. von Hoerner, S. 1955, ZAp, 35, 255
457. Walker, A. R. 1989, AJ, 98, 2086
458. Walker, G. 1987, Astronomical Observations (Cambridge University Press)
459. Waller, W. H. 1991, ApJ, 370, 144
460. Wang, Z., Schweizer, F., & Scoville, N. Z. 1992, ApJ, 396, 510
461. Watson, A. M. 1996, AJ, 112, 534
462. Webb, T. 1998, MSc Thesis, McMaster University
463. Weil, M. L., & Hernquist, L. 1996, ApJ, 460, 101
464. West, M. J., Côté, P., Jones, C., Forman, W., & Marzke, R. O. 1995, ApJ, 453, L77
465. White, R. E. 1987, MNRAS, 227, 185
466. Whitmore, B. C. 1997, in The Extragalactic Distance Scale, ed. M. Livio, M. Donahue, & N. Panagia (Baltimore: StScI), 254
467. Whitmore, B. C., Gilmore, D. M., & Jones, C. 1993, ApJ, 407, 489
468. Whitmore, B. C., Miller, B. W., Schweizer, F., & Fall, M. 1997 AJ, 114, 797
469. Whitmore, B. C., & Schweizer, F. 1995, AJ, 109, 960
470. Whitmore, B. C., Sparks, W. B., Lucas, R. A., Macchetto, F. D., & Biretta, J. A. 1995, ApJ, 454, L73
471. Wiklind, T., Combes, F., & Henkel, C. 1995, AAp, 297, 643
472. Woodworth, S., & Harris, W. E. 1999, in preparation
473. Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C., & Aaronson, M. 1989, ApJ, 345, 759
474. Zepf, S. E., & Ashman, K. M. 1993, MNRAS, 264, 611
475. Zepf, S. E., Ashman, K. M., English, J., Freeman, K. C., & Sharples, R. M. 1999, AJ, in press (astro-ph/9904247)
476. Zepf, S. E., Ashman, K. M., & Geisler, D. 1995, ApJ, 443, 570
477. Zinn, R. 1980, ApJ, 241, 602
478. Zinn, R. 1985, ApJ, 293, 424
479. Zinn, R. 1993a, in The Globular Cluster – Galaxy Connection, ASP Conf. Series 48, edited by G. Smith & J. P. Brodie (San Francisco: ASP), 38
480. Zinn, R. 1993b, in The Globular Cluster – Galaxy Connection, ASP Conf. Series 48, edited by G. Smith & J. P. Brodie (San Francisco: ASP), 302
481. Zinn, R. 1996, in Formation of the Galactic Halo ... Inside and Out, ASP Conf. Series 92, edited by H. Morrison & A. Sarajedini (San Francisco: ASP), 211
482. Zinn, R., & West, M. J. 1984, ApJS, 55, 45
483. Zinnecker, H., McCaughrean, M. J., & Wilking, B. A. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson: Univ.Arizona Press), 429