THE ASTRONOMICAL JOURNAL, 121:2584–2596, 2001
© 2001. The American Astronomical Society. All rights reserved. Printed in U.S.A.

HUBBLE SPACE TELESCOPE-NICMOS OBSERVATIONS OF M31's METAL-RICH GLOBULAR CLUSTERS AND THEIR SURROUNDING FIELDS. I. TECHNIQUES[1]

ANDREW W. STEPHENS AND JAY A. FROGEL
Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210
WENDY FREEDMAN AND CARME GALLART
Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101
PASCALE JABLONKA
Observatoire de Paris, Place Jules Janssen, F-92195 Meudon Cedex, France
SERGIO ORTOLANI
Osservatorio Astrofisico di Asiago, Università di Padova, via dell'Osservatorio, I-36012 Asiago, Italy
ALVIO RENZINI
European Southern Observatory Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany
R. MICHAEL RICH
Department of Physics and Astronomy, University of California at Los Angeles, Los Angeles, CA 90095
AND
ROGER DAVIES
Department of Physics, University of Durham, South Road, Durham DH1 3LE, England, UK
Received 2000 November 2; accepted 2001 January 22

ABSTRACT

Astronomers are always anxious to push their observations to the limit—basing results on objects at the detection threshold, spectral features barely stronger than the noise, or photometry in very crowded regions. In this paper we present a careful analysis of photometry in crowded regions and show how image blending affects the results and interpretation of such data. Although this analysis is specifically for our NICMOS observations in M31, the techniques we develop can be applied to any imaging data taken in crowded fields; we show how the effects of image blending will limit even the *Next Generation Space Telescope*. We have obtained *HST*-NICMOS observations of five of M31's most metal-rich globular clusters. These data allow photometry of individual stars in the clusters and their surrounding fields. However, to achieve our goals—obtain accurate luminosity functions to compare with their Galactic counterparts, determine metallicities from the slope of the giant branch, identify long-period variables, and estimate ages from the AGB tip luminosity—we must be able to disentangle the true properties of the population from the observational effects associated with measurements made in very crowded fields. We thus use three different techniques to analyze the effects of crowding on our data, including the insertion of artificial stars (traditional completeness tests) and the creation of completely artificial clusters. These computer simulations are used to derive threshold- and critical-blending radii for each cluster, which determine how close to the cluster center reliable photometry can be achieved. The simulations also allow us to quantify and correct for the effects of blending on the slope and width of the RGB at different surface brightness levels. We then use these results to estimate the limits blending will place on future space-based observations.

Key words: galaxies: individual (M31) — galaxies: star clusters — techniques: photometric

1. INTRODUCTION

The main objective of our observations is to obtain physical parameters for a selection of metal-rich globular clusters in M31. These parameters are usually straightforward to derive from color-magnitude diagrams (CMDs), however, we have found the effects of crowding to be particularly severe. We therefore present our analysis in two parts. This paper describes in detail the effects of crowding, how we quantify them, the techniques used to correct for them, and their implications for future space-based observations such as with the *Next Generation Space Telescope* (*NGST*). A second paper (Stephens et al. 2001, hereafter Paper II) presents the science obtained from our observations using these techniques.

The general effects of crowding are well known, although not always accounted for. These effects include reduced photometric accuracy, which broadens the CMD and smears the luminosity function (LF), reduced positional accuracy (Hogg 2001), artificial brightening and shifts in measured colors (Aparicio & Gallart 1995), and in severe cases the creation of objects brighter than anything in the parent population through random grouping and blending of many stars (Renzini 1998; DePoy et al. 1993). If these effects are not taken into account, they can result in a mistaken spread in metallicity assumed from the red giant branch (RGB) width, an incorrect mass function extrapolated from the LF, or a false age estimation from the asymptotic giant branch (AGB) peak luminosity.

[1] Based on observations with the NASA/ESA *Hubble Space Telescope* obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy (AURA), Inc., for NASA under contract NAS 5-26555.

This paper is organized as follows. Section 2 presents the details of our observations, and § 3 explains the reduction procedures. Section 4 describes the different simulations in detail. Section 5 presents the results which are applied to the cluster data in Paper II, and § 6 shows how these results can be applied to future space-based observations, such as the *NGST*. We summarize our main conclusions in § 7.

2. OBSERVATIONS

We have obtained *HST* NICMOS images of five of M31's metal-rich globular clusters and their surrounding fields (Cycle 7; program 7826). The clusters are G1, G170, G174, G177, and G280; however most of the discussion in this paper will focus on G280, which has the densest core, one of the sparsest surrounding fields, and is dispersed enough to allow accurate cluster star measurements in its outer regions.

Our observations were taken with the NICMOS camera 2 (NIC2) which has a plate scale of $\sim 0\farcs0757$ pixel^{-1} and a field of view of $19\farcs4$ on a side (376 arcsec^2). The NICMOS focus was set at the compromise position 1-2, which optimizes the focus for simultaneous observations with cameras 1 and 2. All of our observations used the MULTIACCUM mode (MacKenty et al. 1997) because of its optimization of the detector's dynamic range and cosmic-ray rejection.

Each of our targets was observed through three filters: F110W (0.8–1.4 μm), F160W (1.4–1.8 μm), and F222M (2.15–2.30 μm). These filters are close to the standard ground-based *J*, *H*, and *K* filters. Each globular cluster was observed over three orbits of *HST*, with $\sim$42 minutes of observations per orbit. This yielded total integration times of 1920 s in F110W, 3328 s in F160W, and 2304 s in F222M (see Table 1).

We implemented a spiral dither pattern with four positions to compensate for imperfections in the infrared array. The dither steps were $0\farcs4$ for the *J*- and *K*-band images, and $5\farcs0$ for the *H*-band images. Thus the full size of the combined dithered images is $\sim 20''$ in *J* and *K* and $\sim 24''$ in *H*.

TABLE 1

NICMOS FILTERS

FILTER	EXPOSURE (s)	FWHM (pixels)	FWHM (arcsec)
F110W.....	1920	1.65	0.13
F160W.....	3328	1.95	0.15
F222M.....	2304	2.45	0.19

We used the predefined sample sequences STEP32 with 22 samples in *J* and 25 samples in *K*, and STEP64 with 21 samples in *H* (MacKenty et al. 1997).

The *H*-band image of G280 is shown in the left side of Figure 1. This is the combination of four dither positions and is $\sim 24''$ on a side.

3. DATA REDUCTION

Our data were reduced with the STScI pipeline supplemented by the Image Reduction and Analysis Facility (IRAF[2]) NICPROTO package (1999 May) to eliminate any residual bias (the "pedestal" effect). To account for any bias shift which may have occurred in the NICMOS reduction (arbitrary amounts get added or subtracted during pedestal removal), or unsubtracted background contribution, we scale our surface brightness by adding a constant to the entire frame to agree with the surface brightness from Kent (1987). As an example, the procedure for the G280 frame was as follows. Taking the observed position of G280 as $20\farcm5$ from the center of M31 and 34° from the major axis, we

[2] IRAF is distributed by the National Optical Astronomy Observatories, which are operated by AURA, under cooperative agreement with the National Science Foundation.

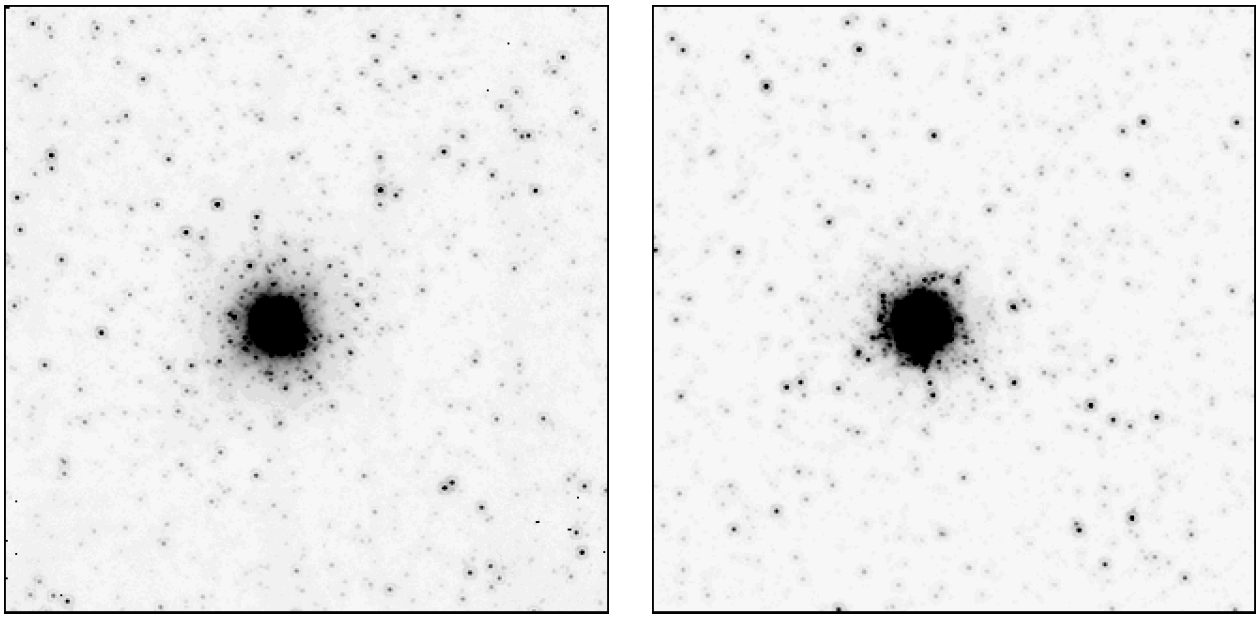


FIG. 1.—*Left*: G280—F160W (*H*-band) combination of all four dithers; 3328 s total exposure. *Right*: A280—Artificial analog of the G280 frame, composed of 450,000 cluster stars and 80,000 field stars.

interpolate Kent's data to find an equivalent major axis distance of ~23.'8 from M31's center. At this distance the total r-band surface brightness is $\mu_r = 21.2$ mag arcsec^{-2}. Assuming an $(r-K)$ color of 2.9 from Terndrup et al. (1994) for the disk of a galaxy similar to M31, we estimate the K-band surface brightness to be 18.3 mag arcsec^{-2}. We then added a constant 0.045 to the count rate in each pixel of our frame to bring the average surface brightness of the frame up to this level (measured greater than ~8″ around the cluster).

Object detection was performed with DAOFIND on a combined image made up of all the dithers of all the bands (12 images in total). Point-spread functions (PSFs) were constructed with DAOPHOT for each of the four dithers, then averaged together to create a single PSF for each band of each target. We use a model PSF with a radius of 8 pixels (0.''6) which is ~3–5 times the stellar full-width at half-maximum (FWHM; Table 1). Instrumental magnitudes were measured using the ALLFRAME PSF-fitting software package (Stetson 1994), which simultaneously fits PSFs to all stars on all dithers. DAOGROW (Stetson 1990) was used to determine the best magnitude in a 0.''5 radius aperture, which we then converted to the CIT/CTIO system using the transformation equations of Stephens et al. (2000).

4. ARTIFICIAL STAR TESTS

The measurements of stars in the M31 clusters and their surrounding fields will be affected by crowding due to the high stellar density. We have performed three types of tests to evaluate the effects of crowding on our photometry and to quantify at what surface brightness levels we can achieve accurate measurements. We compare these techniques with one another (§ 4.3) and with predictions from first principles (§ 5.2).

4.1. *Traditional Completeness Tests*

Our first attempt to determine the effects of crowding, as well as to quantify our photometric completeness, used traditional artificial star tests (Stetson & Harris 1988). The philosophy is that, by "secretly" inserting just a few stars, when the reductions are complete, we a priori know the true brightnesses and colors of these stars so that we can estimate our completeness and errors.

We carried out this test by adding a small number (81) of stars to every raw image. These 81 stars were arranged in a grid with spacing of $2 \times R_{\rm PSF} + 2 = 18$ pixels to avoid self-crowding. The stars were added with ADDSTAR (Stetson 1987) including random Poisson noise, and all had the same integer instrumental magnitude. In order to insure good spatial coverage, the grid of input stars was shifted by 3 pixels and the process repeated. Thirty-six trials and 2916 artificial stars later, we had added an artificial star to every ninth pixel in the region of the cluster.

After the stars were added, we analyzed these synthetic frames in exactly the same way as our original frames. We detected stars on a combined image using DAOFIND and measured each set of dithers simultaneously with ALLFRAME. Once the final photometry list was completed, we extracted the colors and magnitudes recovered for the input artificial stars using DAOMASTER to match recovered coordinates with input coordinates. Since we were attempting to perform a blind test to understand the effects of completeness and blending, we only used coordinates to match input and recovered stars. We could have required that the recovered star be nearly the same brightness as the input star, but at the cost of precluding measuring blending.

We repeated this 36-trial procedure for seven different integer instrumental input magnitudes. After the photometric calibration the input stars have $M_K = $ n.3 where n is an integer ranging from -2 to -8, with fixed colors of $(J-H) = 0.3$ and $(J-K) = 1.6$. The recovered magnitudes for three of the seven sets of trials (input magnitudes of $M_K = -7.3, -5.3, -3.3$) are shown in Figure 2.

Figure 3 shows a radial plot of the G280 frame. Points represent real objects measured on the original frame, and we have overplotted the loci of the recovered artificial stars. These loci are polynomial fits to the recovered artificial stars for each input magnitude (see Fig. 2). The plotted loci are for $M_K = -3.3$ to $M_K = -8.3$; the artificial star set with $M_K = -2.3$ recovered too few stars to fit an accurate locus. At large radii (~5″) there is good correspondence between the input and the recovered magnitudes; thus the actual value of each set of input magnitudes can be obtained from the level of each line at $r > 5''$. However, at small radii, what is measured can be several magnitudes above what was input. Regardless of the star's input brightness, if a star is recovered near the center of the cluster, it is recovered with the luminosity of the object on which it fell; if it does not fall near a bright star, then it is most likely not recovered.

Note that the brightest objects measured on the real frame are all within the central few arcseconds and that the upward curving loci of the artificial stars closely mimic the distribution of those bright stars measured on the real frame. This is indicative of the blended nature of these objects but does not reveal their true luminosities, as *all* the

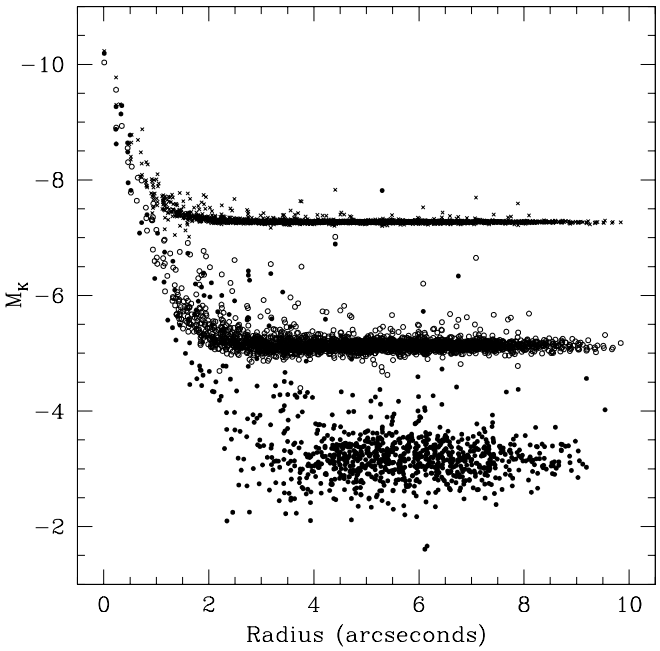


FIG. 2.—Recovered traditional artificial star magnitudes for three input magnitude trials as a function of distance from the center of the G280 cluster. The upper set of points indicate stars input at $M_K = -7.3$ (*crosses*), the middle set were input at $M_K = -5.3$ (*open circles*), and the lower set at $M_K = -3.3$ (*dots*). Each magnitude set represents the results of 2916 artificial stars input in 36 individual trials.

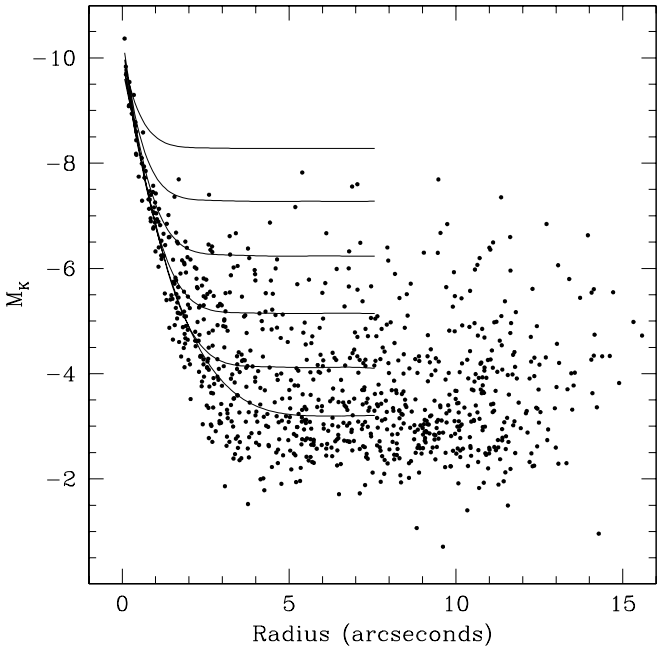


FIG. 3.—Stars measured on the observed G280 frame (*points*), and the loci of the *recovered* artificial stars for each input magnitude from the traditional completeness tests. The plotted artificial stars were input with $M_K = n.3$, where n is an integer from -3 to -8. The top line is the locus of measurements resulting from stars input with $M_K = -8.3$, and the bottom line is from $M_K = -3.3$. The upward curving of these lines results from the increasing difference between input and recovered magnitudes at decreasing radii.

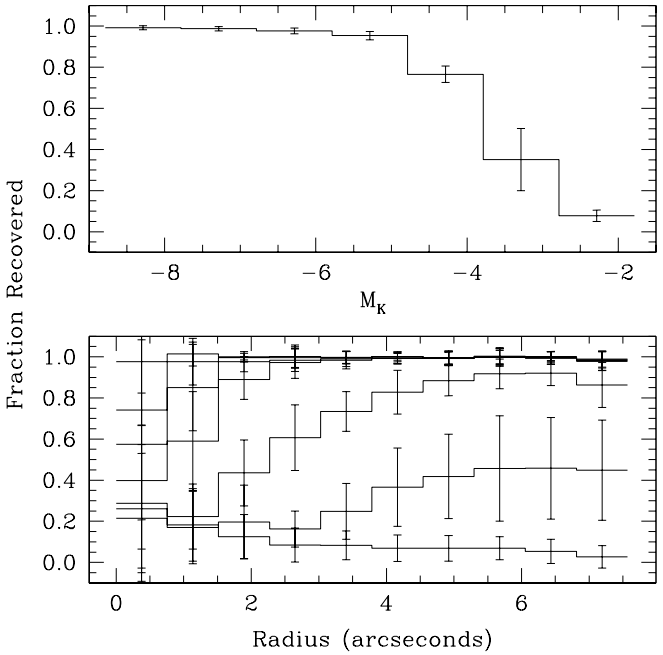


FIG. 4.—*Top panel*: Fractional completeness of the G280 observations as a function of input magnitude for the region sampled by the artificial star tests. *Bottom panel*: Completeness as a function of distance from the center of the cluster for each input magnitude. Each bin in the top panel has its own histogram in the lower panel, thus the lowest histogram in the bottom panel corresponds to the faintest bin in the top panel ($M_K = -2.3$), and the most complete histogram to the brightest bin in the top panel ($M_K = -8.3$).

loci of the artificial stars pass through the brightest measured objects at the center of the cluster.

Figure 4 shows the completeness estimated for each input magnitude by comparing the numbers of stars input and recovered. The top panel shows the fractional completeness for the entire region of the G280 frame sampled by the artificial star grid (13.″4 × 13.″4) as a function of input magnitude. Each bin represents a total of 2916 artificial stars added 81 stars at a time; the error bars show the dispersion in the number of recovered stars over the 36 trials. This plot shows that over the region sampled, we are on average nearly 100% complete down to $M_K \sim -5$, dropping to only 10% complete at $M_K \sim -2$.

The bottom panel of Figure 4 shows the completeness of each magnitude set as a function of position from the center of the cluster. Each bin in the upper panel has its own histogram in the lower panel. The most incomplete histogram corresponds to $M_K = -2.3$, and the most complete histogram to $M_K = -8.3$.

Since we are not using a delta-magnitude criterion when matching input and recovered stars, we expect that misidentifications will increase the completeness estimate near the cluster center where the density of stars is highest. This shift can be seen in the lowest ($M_K = -2.3$) histogram in the bottom panel of Figure 4, where the completeness actually increases near the center of the cluster.

However, this type of simple completeness histogram does not show one of the most dangerous effects of crowding: bin hopping. Bin hopping is when stars are input in one magnitude bin but are recovered in another, usually brighter bin. This effect can be clearly seen at small radii in Figure 3, where regardless of which input locus you examine, when the surface brightness (crowding) gets high enough, those input stars are on average recovered significantly brighter than their original input magnitude.

4.2. *Artificial Clusters*

4.2.1. *Generation of the Artificial Clusters and Fields*

From the photometric results in Figure 3, we see that we measure predominantly very bright objects in crowded regions. The artificial star tests showed that the measurement of faint stars injected into these crowded regions will be significantly altered by the severe crowding, and these faint stars will usually be absorbed into the nearest bright object. To further investigate the nature of these bright objects, we decided that more tests were necessary. However, heeding the warning of DePoy et al. (1993), that typical artificial star experiments cannot reproduce the true effects of severe image crowding, and, following the lead of Rich, Mould, & Graham (1993), we initiated our own set of "nonstandard" artificial star experiments.

These experiments involved the simulation of the entire cluster and surrounding field using a reasonable estimate of the expected cluster star properties. Each artificial dither starts as a blank frame with the appropriate amount of random noise. To this we randomly add cluster and field stars. We use PSFs determined from the G280 cluster with any negative values in the PSF set to zero and include random Poisson noise in each. We approximate the luminosity function and colors of the cluster and field stars with those observed in the Galactic bulge. The LF is a power law with a slope of 0.278, extending from $-7.2 < M_K < 5.8$,

which corresponds to $7.0 < K < 20$ at Baade's Window (Frogel & Whitford 1987; Tiede, Frogel, & Terndrup 1995; DePoy et al. 1993). Colors are derived from observations for $M_K < 2.8$ (Tiede et al. 1995) and from stellar models for $M_K > 2.8$ (Cassisi et al. 2001). Cluster stars are added according to the radial surface brightness profile measured in the real cluster until the integrated magnitude of the cluster matches that of the cluster being modeled. Note that the measured surface brightness is independent of crowding, and as long as there are no saturated pixels, which there were not, the measured surface brightness is accurate. Field stars are added randomly, following the same colors and LF as the cluster stars, to approximately match the observed field star density. The artificial frames are then processed and measured in exactly the same manner as the real data, namely finding stars on a combined image with DAOFIND, then measuring all dithers simultaneously with ALLFRAME.

4.2.2. *The Artificial Cluster CMD*

In order to best understand the effects of crowding, we focus on the G280 observations. G280 has one of the densest cores and is one of the most extended clusters in our sample. This cluster also has one of the sparsest fields of any of the clusters observed and hence the largest contrast in surface brightness between field and cluster. The artificial analog of the G280 cluster, which we call A280, is shown on the right side of Figure 1. This frame is composed of 450,000 cluster stars and 80,000 field stars, extending over $-7.2 < M_K < 5.8$.

Figure 5 shows the input (*dashed*) and measured (*solid*) luminosity functions for the A280 frame. The top panel is for the whole frame. Even though the input LF cuts off at $M_K = -7.2$, the measured LF extends to $M_K \sim -10$. Based on the distance where the cluster blends into the background, we choose a radius of 5″ to divide the frame into cluster and field regions. The middle panel shows the LF for the field around G280, only counting objects located greater than 5″ from the cluster center. Here the measured LF shows very good correspondence with the input LF with identical bright end cutoffs. The faint end decline at $M_K \sim -2$ is due to incompleteness. The bottom panel includes only objects closer then 5″ from the cluster center; we assume that these are mostly cluster stars. Here is where the major discrepancy arises between the input and recovered LFs. All of the objects measured brighter than the input LF cutoff, i.e., brighter than any single star, are found in this small region of the frame. We emphasize that the same input LF was used for the entire frame.

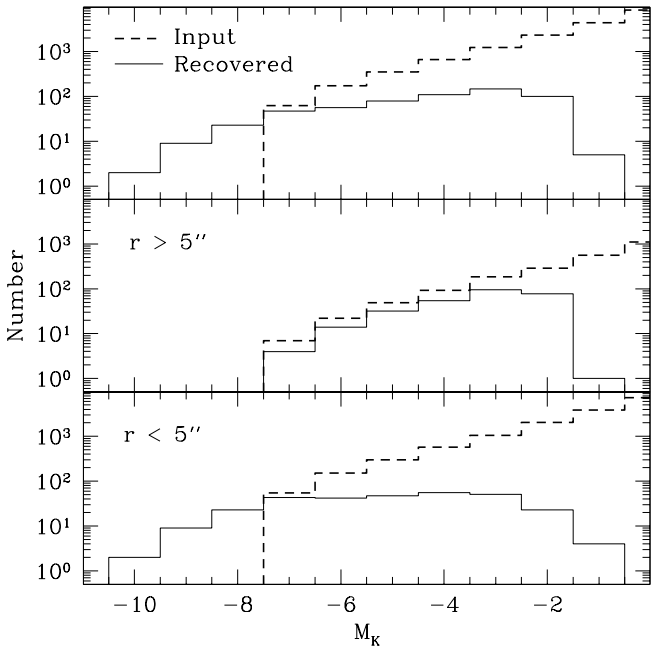


FIG. 5.—Artificial cluster A280 input (*dashed*) and recovered (*solid*) luminosity functions. *Top*: Entire frame. *Middle*: Only stars located >5″ from the cluster center. *Bottom*: Stars found <5″ from the cluster center. The faint end decline is due to incompleteness, the extension of the measured LF brighter than the input LF is due to blending.

The A280 CMDs are shown on the right side of Figure 6 where we have again divided the field spatially. Objects closer than 5″ from the cluster center are plotted in the left panel, while objects farther than 5″ are in the right panel. Points and circles indicate measured objects; the locus of the input stars is indicated by the line (the RGB of the Galactic bulge). In the $r > 5''$ panel there is a very good correspondence between the input and recovered magnitudes. There is very little crowding and as a result very little scatter of the measured data around the input data. The amount of scatter at different brightnesses gives an idea of the scale of the measurement errors. In the $r < 5''$ panel there is a large plume of stars measured brighter than anything input into the cluster and bluer than the RGB tip. This blue color is an immediate indicator of the blended origin of these objects, where what is measured is not a single bright star, but the superposition of many fainter, bluer stars. Using criteria developed in § 5, we plot objects inside the threshold-blending radius, where the blending is just beginning to affect the faintest stars, at surface brightness of $\mu_K = 16$ mag arcsec^{-2}, with open circles. Objects measured inside the critical-blending radius, where the surface brightness is $\mu_K \leq 14$, and blending has severely distorted nearly all measurements, are indicated with half-size dots. Note further that nearly all recovered stars lie systematically to the blue of the input line for $r < 5''$. This suggests that blending is affecting all stars to some degree, which is not surprising since $r \sim 5''$ is also the dividing line inside which we added more than one star per pixel in order to match the observed brightness profile with a fixed LF.

The G280 CMDs are shown on the left of Figure 6 for comparison. The left panel shows all of the data inside a radius of 5″, the distance chosen to define the cluster. Objects inside the threshold-blending radius ($r < 2\farcs2, \mu_K < 16$) are plotted with open circles, and objects inside the critical-blending radius ($r < 1\farcs0, \mu_K < 14$) are plotted with half-size dots. These radii are determined in § 5 and listed in Table 4. Just as in the artificial cluster, there is a blue plume of stars brighter than anything seen in the surrounding field. However, these objects are all found within the critical blending radius. Remember that stars on the RGB will be the reddest stars in a cluster, except for very fainter lower main sequence stars. The right panel shows all objects outside the 5″ cluster radius. These stars are expected to be noncluster, or "field," stars. Open circles indicate objects measured in the high-background region (bottom 25 pixels) of the *K*-band images. The color spread in the G280 field CMD is real. This field is ∼85% disk and ∼15% bulge, with an estimated spread in [Fe/H] of ∼0.4 (Paper II). The artificial A280 frame, in contrast, was created using a single metallicity and thus shows a much smaller color spread.

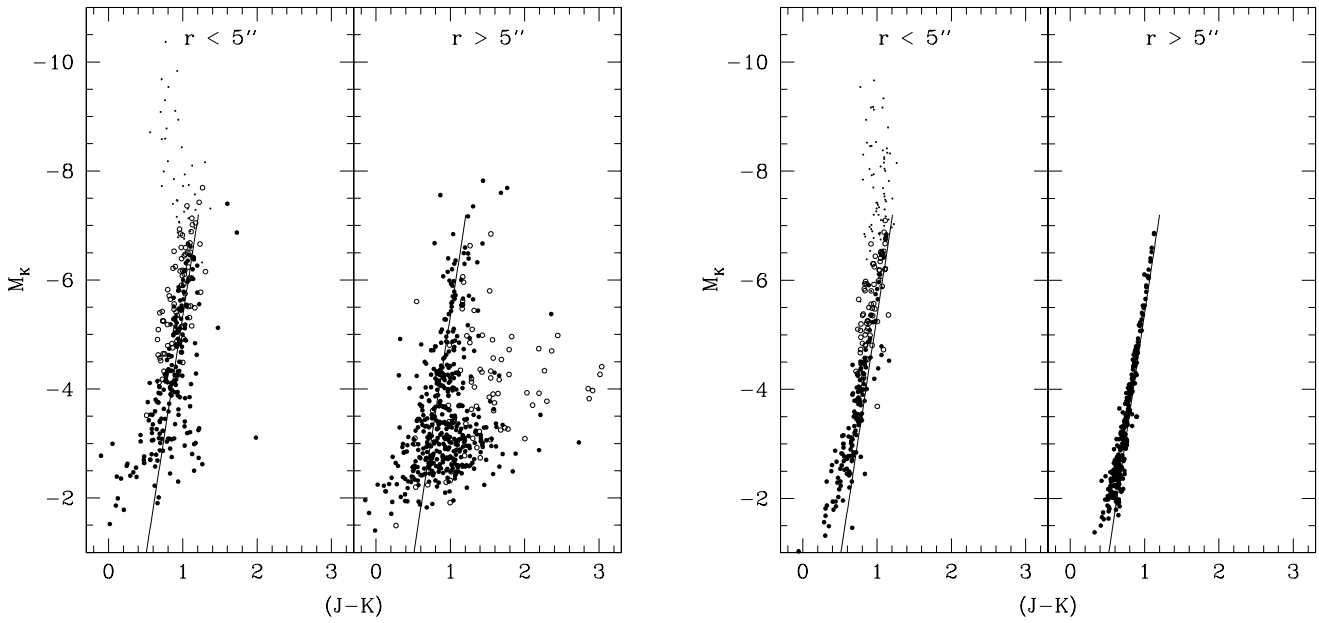


FIG. 6.—CMDs of G280 (*left figure*) and its artificial analog A280 (*right figure*). The left panel of each figure shows objects measured within 5″ of the cluster center, and the right panels objects farther than 5″. Open circles in the left panels indicate objects found within the threshold-blending limit where $\mu_K < 16$ mag. arcsec^{-2}. Open circles in the right panel of the left figure mark objects measured in a region of high background in the bottom 25 pixels of the real K-band images. Half-size dots designate objects inside the critical-blending limit where $\mu_K < 14$ mag arcsec^{-2}. The solid line is the RGB of the Galactic bulge, which is the input into the artificial cluster.

The field subtracted cluster CMD of G280 also looks remarkably similar to the A280 cluster CMD (Paper II).

Table 2 summarizes the average blending of stars over a range in measured luminosities in the A280 cluster. To construct this table, we have selected the five stars closest to each integer recovered magnitudes. At each magnitude, we average these five star's radial distance from the cluster center and the number of neighbors which are contributing to their measured flux. The first column lists the average recovered K-band magnitude of the selected stars. Columns (2) and (3) give the average distance of the stars from the center of the cluster and the average surface brightness at that distance. The fourth column lists the average number of stars brighter than $M_K = -2$ which were input within a 2 pixel radius of the measured star's position (0.″15; about the size of one resolution element). The last column is the average number of stars input within 2 pixels of the measured star's position at any magnitude.

Table 2 shows why objects measured near the cluster center are measured so bright. The five stars measured nearest $M_K = -9$, all of which lie approximately 0.″2 from the center of the cluster, have on average 150 neighbors brighter than $M_K = -2$ that are close enough to contribute to their measured flux. The tight relationship between measured brightness and cluster distance continues down to $M_K \sim -7$, where the average distance is $\sim 1''$ and $\mu_K \sim 14$. Stars measured fainter than this have a much larger range in recovered position and also have only a few "bright" stars ($M_K < -2$) near enough to distort their measured brightnesses.

TABLE 2

A280 AVERAGE BLENDS[a]

$\langle M_K \rangle$ (1)	⟨Radius⟩ (arcsec) (2)	μ_K (3)	$\langle N_{2pix} \rangle$ ($M_K < -2$) (4)	$\langle N_{2pix} \rangle$ ($M_K < 5.8$) (5)
-9.12 ± 0.10...	0.2 ± 0.1	11.3	151 ± 49	23769 ± 7772
-7.99 ± 0.06...	0.7 ± 0.1	13.2	37 ± 7	5315 ± 1351
-7.00 ± 0.03...	0.9 ± 0.1	13.7	17 ± 8	2675 ± 944
-6.00 ± 0.03...	2.8 ± 2.0	16.5	3.4 ± 3.8	577 ± 573
-4.96 ± 0.03...	3.1 ± 2.1	16.8	2.8 ± 2.6	301 ± 288
-4.00 ± 0.00...	6.6 ± 4.0	18.4	0.0 ± 0.0	19 ± 15
-3.01 ± 0.01...	6.5 ± 4.4	18.5	0.2 ± 0.4	23 ± 23

[a] Average of five stars at each magnitude taken from one simulation.

4.2.3. *Effects of Varying the Input LF*

We have run simulations varying the shape of the LF input into the artificial cluster. The results are summarized in Table 3. The first two columns list the limits of the power-law input LF, and the third column is the power-law slope α [$N(m) \propto 10^{\alpha m}$]. The fourth column gives the number of stars input into the cluster; this number is scaled to keep the integrated magnitude of the cluster constant. The last two

TABLE 3

A280 VARIABLE LFS

INPUT LIMITS					
M_{K1}	M_{K2}	α	N_{input}	M_K[a] (brightest)	$(J-K)$[b] (top five)
5.8.....	−9.2	0.278	254,000	−9.85	1.11
5.8.....	−8.2	0.278	338,000	−9.92	1.03
5.8.....	−7.2	0.278	450,000	-9.70 ± 0.13	0.96 ± 0.01
5.8.....	−6.2	0.278	601,000	−9.93	0.87
5.8.....	−5.2	0.278	805,000	−9.41	0.72
5.8.....	−7.2	0.238	207,000	−9.74	1.12
5.8.....	−7.2	0.258	306,000	−9.77	0.93
5.8.....	−7.2	0.278	450,000	-9.70 ± 0.13	0.96 ± 0.01
5.8.....	−7.2	0.298	653,000	−9.61	0.93
5.8.....	−7.2	0.318	933,000	−9.69	0.93

[a] Brightest recovered star.
[b] Average $(J-K)$ color of the brightest five recovered stars.

columns list the brightest star measured in each cluster and the average $(J-K)$ color of the five brightest stars in each cluster.

For reference, the "normal" A280 LF taken from Baade's Window has a slope of $\alpha = 0.278$ extending from $-7.2 < M_K < 5.8$ and was constructed from 450,000 stars. We ran this normal simulation five times using different random number seeds in order to asses the range of potential results for a fixed input LF. These five simulations yielded a mean M_K(brightest) of -9.70 ± 0.13 and $(J-K)_{\text{Top5}} = 0.96 \pm 0.01$.

The top half of Table 3 describes the simulations where we varied the bright end extent of the input LF by ± 2 mag. All the simulations have the same integrated cluster magnitude and achieve approximately the same core surface brightness. These simulations also generate blends which reach nearly the same brightness ($\langle M_K(\text{brightest})\rangle = -9.75 \pm 0.22$) with a scatter not much larger than that observed in five simulations using the same input LF ($\sigma = 0.13$). The most noticeable difference between these simulations is the recovered colors. The simulation with the faintest input LF cutoff and hence the most stars, returns objects with bluer colors. At the top of the CMD, the stars from the faintest input LF cutoff simulation have $(J-K) \sim 0.4$ mag bluer than compared with the simulation with the brightest input LF cutoff.

The bottom half of Table 3 lists the simulations where we varied the input LF slope by ± 0.04. As before, all simulations achieve approximately the same core surface brightness and brightest blends. The average M_K(brightest) is -9.69 ± 0.06 and has a 1 σ dispersion which is actually smaller than that seen in the five runs using the same "normal" LF. The simulations with a steeper input LF have a noticeably wider RGB due to the many blends of fainter stars. However, the brightest blends do not show the same gradual progression toward bluer colors with increased number of input stars as was observed when varying the input LF cutoff.

Can these simulations recover the true LF from a blended frame? In the fields surrounding the clusters, the simulations confirm that what we measure is indeed the true LF. However, in the cores of the clusters where the stars are severely blended, there is a degeneracy between the LF extent, slope, and number of stars (surface brightness). In these dense regions, using the current observations and techniques, it is nearly impossible to distinguish between a cluster with 250,000 stars as bright at $M_K \sim -9$ and another cluster with 3 times as many stars that only get as bright as $M_K \sim -5$.

4.3. *Comparison of Artificial Star Tests*

We now compare the luminosity functions from the traditional completeness tests, where a handful of artificial stars were added to an observed frame, with those from the artificial cluster tests.

We start with the traditional completeness test LF shown in the top panel of Figure 7. The dashed line illustrates the input LF composed of 2916 input stars per magnitude, constructed from 81 stars at a time over 36 trials. The solid line is the recovered LF. To mimic a power-law input LF, we multiply the input LF by the same power-law slope as was used in the artificial cluster tests ($\alpha = 0.278$). The number of stars recovered (at any magnitude) is then multiplied by the same constant that was applied to the responsible input

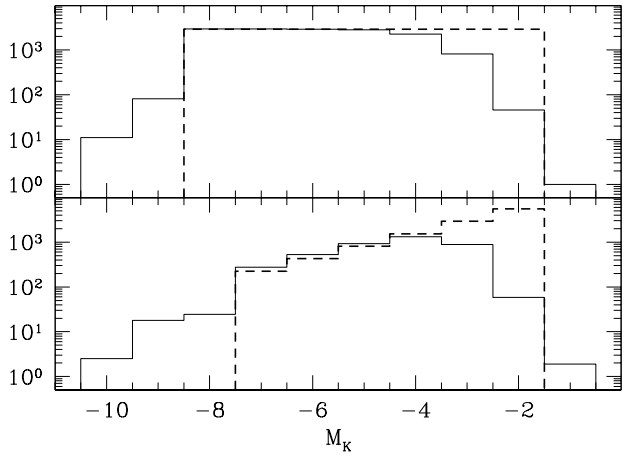


FIG. 7.—Transformation of the traditional completeness test results to mimic the power-law input LF used in the artificial cluster tests. *Top*: Initial boxcar input LF (*dashed*) and the recovered LF (*solid*). *Bottom*: Input LF multiplied by a power-law and truncated at $M_K = -7.3$ (plotted with 1 mag wide bins) and the resulting recovered LF. Note that all stars measured with $M_K < -7.3$ are blends.

magnitude. The recovered stars from each input magnitude are then summed to get the total recovered LF. However, from our initial completeness tests, we input stars brighter than the true extent of the LF. (This is because with just the traditional completeness tests, the true brightnesses of the measured stars are unknown, because one is just placing artificial stars on the preexisting clumps of stars.) Thus we now truncate the input LF at $M_K = -7.3$ to match the artificial cluster input LF as closely as possible. The resulting input (*dashed*) and recovered (*solid*) LFs are shown in the bottom panel of Figure 7.

The traditional completeness test results (*bottom*, Fig. 7) appear very similar to the artificial cluster results (*top*, Fig. 5). The input LFs (*dashed lines*) start at $M_K \sim -7$ and follow a power-law distribution upward toward fainter magnitudes. The observed LF (*solid line*) extends 3 mag brighter than the input LF and drops off at $M_K \gtrsim -2$ because of incompleteness.

At the faint end of the photometry, both tests are equally useful to characterize the photometric completeness. The artificial clusters simulations are, however, a more controlled experiment that offer advantages for the interpretation of the nature of the brightest measured stars. The traditional completeness tests show that some stars are brightened, but when the crowding gets severe, the brightening arises from falling on or near a preexisting bright clump. These tests, however, do not tell you whether the bright clumps are real. The artificial cluster simulations, on the other hand, tell you the exact number and brightness of each of the stars which make up the clumps and, in our case, that these bright objects which are observed can be created with a normal LF whose brightest star is several magnitudes fainter than the brightest object observed.

4.4. *Artificial Fields*

In order to investigate any possible effects of the cluster's strong surface brightness gradient on our blending tests we have created uniformly populated mini-fields of 100×100 pixels ($7\farcs6 \times 7\farcs6$). These mini-fields also allow us to more accurately quantify the effects of blending as a function of surface brightness by providing more recovered stars at

each surface brightness. This is especially important since we use surface brightness as the criterion for quantifying blending and determining the believability of our photometry. Each mini-field is constructed in the same manner as the artificial clusters, using a power-law input luminosity function, extending across $-7.2 < M_K < 5.8$. Between 10^4 and $\sim 5 \times 10^6$ stars are randomly added to each field, where the field with the most stars nearly reaches the surface brightness level observed in the cores of the M31 clusters.

The input (*dashed lines*) and recovered (*solid lines*) luminosity functions of these fields are shown in Figure 8. The surface brightness of each field is listed in the upper right corner of each plot. There is good agreement between the bright end extent of the input and measured LFs for surface brightnesses up to $\mu_K = 15.7$. However as more and more stars are added to the field, blending becomes significant, the measured LF shifts toward brighter and brighter magnitudes, and more stars are "recovered" with M_K brighter than the brightest input star.

Notice the change of the slope of the measured LF. When blending is insignificant, the measured LF is very close to the input LF, i.e., increasing toward fainter magnitudes with the correct slope. However when blending becomes severe, more stars clump together, and fewer faint stars are recovered. The relative number of bright stars compared with faint stars changes dramatically, and the slope of recovered LF now has an opposite sign compared with the input LF.

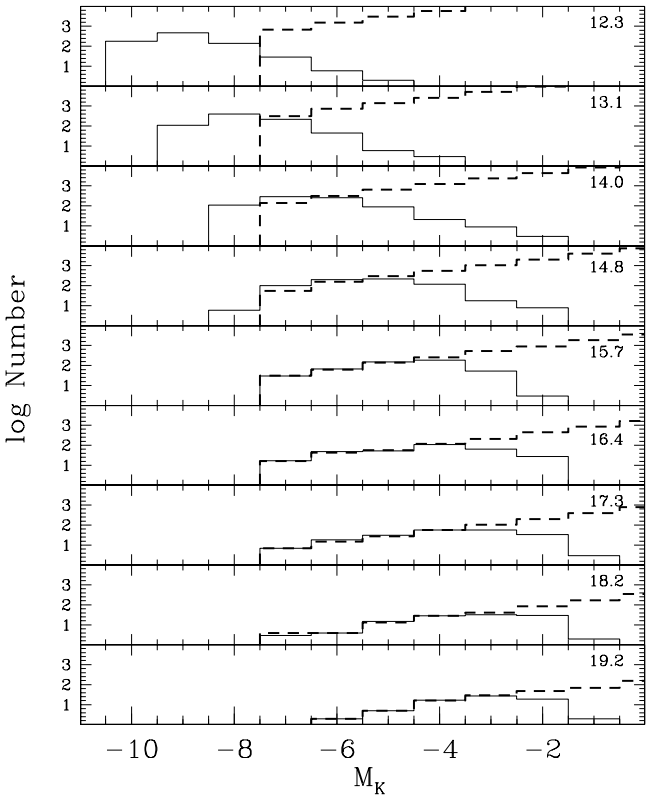


FIG. 8.—Input (*dashed*) and recovered (*solid*) luminosity functions of artificial fields of varying surface brightnesses. All fields have the same input LF slope, and the average surface brightness of each field is listed in the upper right corner of each panel. The brightest star input into any field is $M_K = -7.2$ (binned into 1 mag wide bins), however, because of small number statistics, the less-populated fields may not have any stars this bright. Note how the recovered LF not only shifts toward brighter magnitudes with increasing surface brightness but also changes slope because of the effects of blending.

5. RESULTS

We find that many of the bright objects observed in crowded regions arise from the clumping of many fainter stars. Although the traditional completeness tests do show that faint stars are artificially brightened, they do not give the whole picture. Our artificial frames, which closely mimic the real observations show that random clumps of stars can create objects several magnitudes brighter than the brightest real stars.

The effects of blending can be seen in the left side of Figure 9, a plot of the recovered absolute K-band magnitude as a function of field surface brightness. Here each curve represents the median of stars input with a range of ± 0.5 mag. For example the bottom-most curve gives the median recovered brightness of all stars input with magnitudes between -1.5 and -2.5 in each mini-field. The number of stars recovered in each 1 magnitude range is a function of the total number of stars input (i.e., the surface brightness) and the range being considered, since we use the same power-law input luminosity function. Thus the statistical noise is quite high at the faint surface brightness end where the number of input stars is low, especially for the brighter input curves. Note that there were no stars input with $M_K \sim -6$ or $M_K \sim -7$ at $\mu_K \sim 22$, or with $M_K \sim -7$ at $\mu_K \sim 19$.

The right side of Figure 9 shows a plot of the average deviation between the input and recovered magnitudes for stars of different input brightnesses as a function of the average surface brightness of the artificial field. This plot illustrates how increasing surface brightness has an increasingly strong effect on stellar photometry. At surface brightnesses fainter than $\mu_K \sim 17$ magnitudes arcsecond^{-2} all stars are measured fairly accurately. However at $\mu_K \sim 16$ the fainter stars start to become artificially brightened by crowding. With increasing surface brightness the effects of crowding become more and more significant for brighter stars. At $\mu_K \lesssim 13$ the crowding is so severe that nearly all stars are badly blended or misidentified with random clumps of blended stars, and even the brightest input stars are mismeasured by several magnitudes.

The similarity between the recovered magnitudes on the left side of Figure 9 and the fits to the recovered artificial star magnitudes from the traditional completeness tests shown in Figure 3 is obvious. At some critical surface brightness, or radius in the case of our globular clusters, blending effects begin to dominate. As the surface brightness increases, blending becomes more and more important, until even the brightest stars become lost in the uneven background (for more on the effects of crowding on positional errors see Hogg 2001).

Using Figure 9 as a guide, we chose $\mu_K = 16$ mag arcsec^{-2} as the threshold point where photometry of stars of $M_K \sim -3$ starts to become noticeably affected by blending. Of course this "safe" surface brightness limit would be even fainter if we were concerned with accurate measurements of stars fainter then $M_K \sim -3$. We designate $\mu_K = 14$ as the critical surface brightness where our photometry is dominated by blends. Beyond this level, no measurements are reliable.

Figure 10 shows azimuthally averaged surface brightness profiles of each cluster, where our surface brightness values

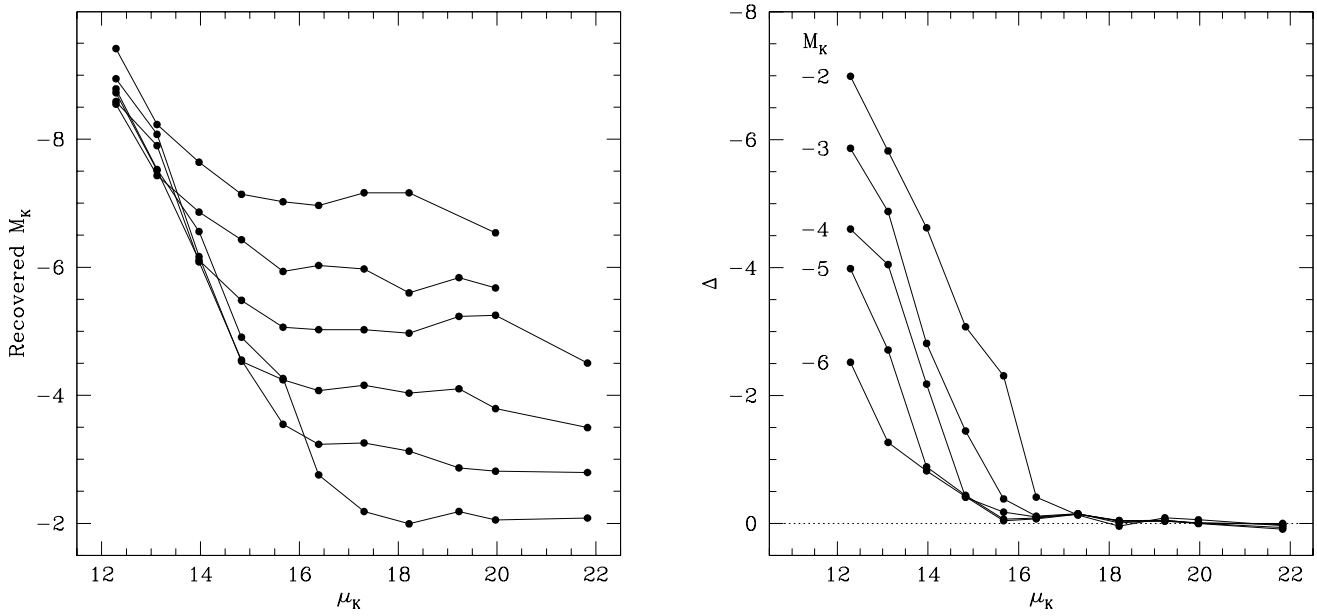


FIG. 9.—*Left*: Recovered magnitude of stars measured in mini-fields over a range of surface brightnesses. Each line indicates the median recovered magnitude over a range of ± 0.5 mag in input brightnesses. The input magnitude can be ascertained from the measurements made at low surface brightnesses, which are accurate but affected by small number statistics. This plot allows one to monitor the effects of crowding on the measurements of stars of approximately constant magnitude. Note the similarity of this plot to Fig. 3. *Right*: Median difference between the input and recovered magnitudes of stars measured in mini-fields of different average surface brightness. Each line indicates a single range of input magnitudes, which are labeled at the leftmost termination of each.

have been normalized using measurements from Kent (1987) as described in § 3. Dotted lines indicate the threshold- and critical-blending surface brightness levels. We use this plot to determine the threshold-blending radius (R_{16}) and the critical-blending radius (R_{14}) for each cluster. These radii are listed in Table 4. Any objects measured inside the threshold-blending radius, especially faint objects, are potentially affected by blending and should be considered suspect. (Remember that the threshold radius R_{16} was chosen so that stars input at $M_K \sim -3$ could be recovered accurately most of the time.) Objects measured inside the critical-blending radius are undoubtedly blends, and, although we plot them for completeness, they should be disregarded.

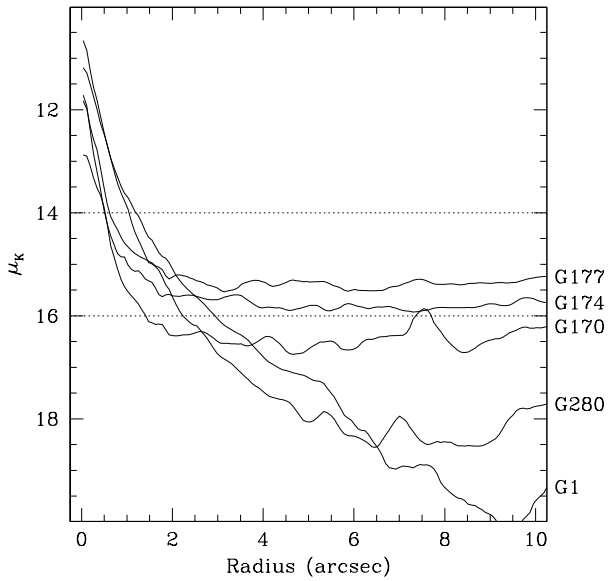


FIG. 10.—Azimuthally averaged K-band surface brightness as a function of radius from the cluster center. The dashed lines show the threshold–($\mu_K = 16$) and critical–blending ($\mu_K = 14$) surface brightness limits.

In summary, traditional completeness tests are good for analyzing one's detection efficiency and photometric errors. However, in the case of blended images, these tests cannot reveal what the blends are composed of. Only by creating completely artificial frames and matching the observed and modeled surface brightnesses can one get an idea of which stellar populations are consistent with the observations. Even then there are degeneracies between the extent of the LF, the slope of the LF, and the true number of stars at high surface brightnesses.

5.1. *Effects of Blending on the RGB Slope and Width*

Since in Paper II we will use the slope of the red giant branch to estimate metallicities and the width of the RGB to place limits on any spread in metallicity, we have investigated the effects of crowding on these quantities. This specific case should serve as a good example for other similar analyses. This analysis uses the mini-fields discussed in the previous section, which include between 10^4 and $\sim 5 \times 10^6$ stars, achieving surface brightnesses in the range $19.2 \lesssim \mu_K \lesssim 12.3$. The input M_K-$(J-K)$ RGB slope $[\Delta(J-K)/\Delta(K)]$ is -0.113, which implies [Fe/H] ~ -0.29 according

TABLE 4

THRESHOLD AND CRITICAL BLEND RADII

Cluster	Core μ_K	R_{14}	R_{16}
G001.....	11.2	1.2	2.9
G170.....	12.9	0.5	1.4
G174.....	11.7	0.5	...
G177.....	11.8	0.6	...
G280.....	10.7	1.0	2.2

to the RGB slope–metallicity relation of Kuchinski & Frogel (1995) and Kuchinski et al. (1995).

We perform an iterative linear least-squares fit to the measured RGB, rejecting measurements farther than $\pm 3\ \sigma$ from the fit. Even though the metallicity relation was derived for the RGB in the range $-1 \gtrsim M_K \gtrsim -6$, here we fit to all the data since blending shifts the RGB several magnitudes in luminosity.

Figure 11 shows the RGB slope measured in each simulated mini-field as a function of the average field surface brightness. The dotted horizontal line at a slope of -0.113 shows the input RGB slope, i.e., the slope which should be recovered in every field if crowding were unimportant. Indeed at the low surface brightness end, the recovered RGB slope approaches the input RGB slope. As the surface brightness of the field increases more faint blue stars clump together pushing the lower end of the RGB bluer and to higher luminosities, thus tilting the RGB and implying higher metallicities. At even higher surface brightnesses ($\mu_K \lesssim 13$) the blending significantly affects even the brightest stars. All faint stars are lost in the background and the brightest stars clumped together with many fainter bluer stars are the only objects measured. The dashed line is a linear fit to the data with $\mu_K > 13$ and is the relation we use to estimate corrections to our measured RGB slopes. We emphasize that this correction is merely an estimate of the effects of crowding. We do not plot the correction as a function of delta slope or delta [Fe/H], in order to discourage blind use, as the effects of blending will be different depending on the specific input luminosity function and colors.

On the right side of Figure 11 is the metallicity scale implied from the relation of Kuchinski & Frogel (1995). This shows the strong dependence of the metallicity on the slope and thus the severe effect that even minimal blending can have on the metallicity derived from the RGB slope.

We now analyze the effects of blending on the width of the measured RGB. All of our simulations start out with an input RGB of zero width. To measure the width of the RGB, the linear fit to the RGB is subtracted off to leave a vertical RGB centered at zero $(J-K)$ color, and the distribution is binned into 0.05 mag color bins. We then fit a Gaussian to the binned distribution. Figure 12 shows the width of the Gaussian fit to the measured giant branch of each field as a function of the field's average surface brightness. This plot shows that the measured width of the RGB is not significantly affected by crowding until the surface brightnesses exceeds $\mu_K \sim 17$. At that point the width grows approximately linearly with increasing surface brightness.

5.2. *Simulations versus Theory*

A great deal on what to expect from crowding can be derived from first principles (Renzini 1998). In this section the main points in this theoretical approach are summarized and then compared with the results of our simulations. For a given stellar population of age t and total bolometric luminosity L_T the number N_j of stars in the generic j (post–main-sequence) evolutionary stage of duration t_j is given by

$$N_{\rm j} = B(t) L_{\rm T}\, t_{\rm j}\ , \tag{1}$$

where $B(t)$ is the specific evolutionary flux, i.e., the number of stars leaving or entering any evolutionary stage per year and per unit (solar) luminosity of the population. In practice, $B(t)$ is only weakly dependent on age, varying from $\sim 0.5 \times 10^{-11}$ to $\sim 2 \times 10^{-11}$ (stars $L_\odot^{-1}$ yr^{-1}, see Figure 1 in Renzini 1998). Equation (1) can be applied to one individual resolution element (re), in which case the sampled luminosity can be written as

$$L_{T,\rm re} = s^2 \Sigma_T = s^2 \Sigma_{T,0} \left(\frac{D}{200\ \rm kpc} \right)^2 , \tag{2}$$

where s is the linear size of the resolution element (in arcseconds), Σ_T is the surface brightness of the stellar system in units of $L_\odot$ arcsec^{-2}, $\Sigma_{T,0}$ is the surface brightness in units of $L_\odot$ pc^{-2}, and D is the distance.

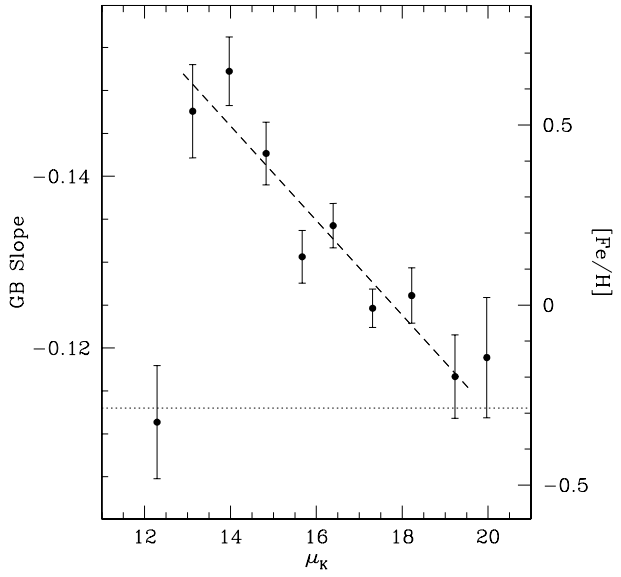


FIG. 11.—Derived giant branch slope and implied metallicity as a function of field surface brightness. The dotted line shows the input RGB slope of -0.113 ([Fe/H] $= -0.29$). The dashed line is a linear fit to data with $\mu_K > 13$.

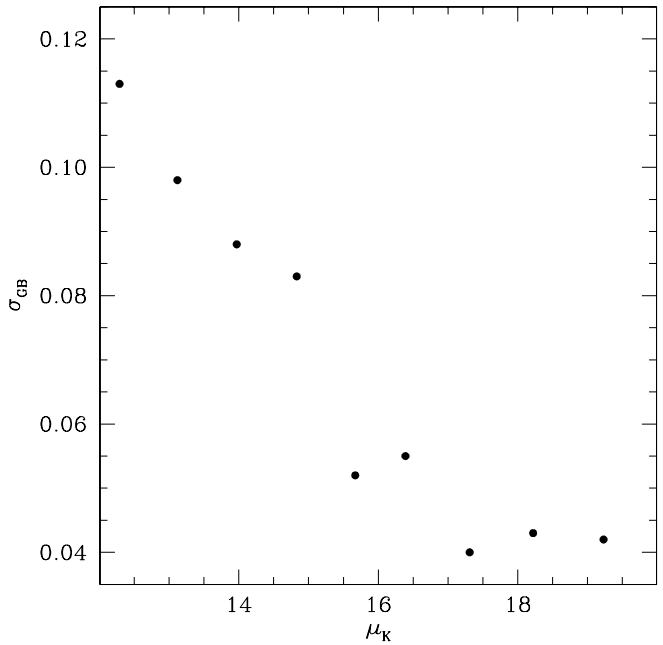


FIG. 12.—Giant branch width as a function of the field surface brightness measured from the artificial mini-fields.

For $N_{j,\mathrm{re}} = B(t)L_{T,\mathrm{re}}t_j < 1$, $N_{j,\mathrm{re}}$ is close to the probability that a generic resolution element contains one star in the j phase. Therefore, $N^2{}_{j,\mathrm{re}}$ is the probability that a resolution element contains a blend of two such stars that a photometric package will mistake for a single star about twice as bright. More generally, the probability that a resolution element will contain a blend of a star in phase j and a star in phase k is given by

$$N_{j,\mathrm{re}}N_{k,\mathrm{re}} = B(t)^2 t_j t_k s^4 \Sigma^2_{T,0}\left(\frac{D}{200\ \mathrm{kpc}}\right)^4. \quad (3)$$

The bottom line is that, for given physical crowding as measured by the surface brightness (in $L_\odot$ pc^{-2} units), the number of two-star blends is proportional to the fourth power of the linear size of the resolution element, to the fourth power of the distance from the observer to the stellar population, and to the square power of the surface brightness (Renzini 1998). For example, suppose we have observed a globular cluster at a distance of 7 kpc with a resolution of 1″ and would like to obtain the same frequency of two-star blends for an identical cluster in M31, i.e., at a distance of 700 kpc. Not surprisingly, equation (3) tells that we would need a resolution 100 times better, i.e., ~0.″01.

As another example, equations (2) and (3) can be used to estimate the average number of RGB stars (N_{RGB}) and RGB stars within 1 mag of the RGB tip (N_{RGBT}) per resolution element. These can then be used to determine the number of resolution elements in a frame that contain a blend of two RGB stars (N_{2RGB}), two RGBT stars (N_{2RGBT}), three RGBT stars (N_{3RGBT}), etc. Note that the number of triplets scales as the cube of N_{RGBT}, that of quadruplets as its fourth power, and so on (Renzini 1998).

The results of our calculations are listed in Table 5. Here we assume a 15 Gyr population of solar metallicity ($L_T = 0.36L_K$), a distance modulus of 24.4, no K-band extinction, a 376 arcsec^2 field of view, and we use the H-band PSF FWHM of 0.″15 as the size of our resolution element.

The first column is the K-band surface brightness (μ_K), and the second is the total bolometric luminosity sampled by a single resolution element (L_T). The third and fourth columns give the average number of stars on the red giant branch (N_{RGB}) and stars within 1 mag of the red giant branch tip (N_{RGBT}) sampled by a single resolution element. The fifth, sixth, and seventh columns give the number of RGBT star blends. N_{2RGBT} is the number of blends of two RGBT stars, N_{3RGBT} is the number of blends of three RGBT stars, and the number of blends of N RGBT stars is given by N_{NRGBT}. The eighth column lists the number of blends of two stars on the RGB (N_{2RGB}) expected on the entire frame. The last column is the total K-band magnitude sampled by one resolution element assuming uniform surface brightness.

The result is graphically shown in Figure 13. The dashed line indicates the predicted number of blends of two RGBT stars (N_{2RGBT}), and the solid line shows blends of two or more RGBT stars (N_{NRGBT}) in the area of one NIC2 frame. Overplotted are the number of stars measured brighter than the maximum input luminosity (L_{max}) in the mini-fields (§ 4.4) normalized to the size of the NIC2 frame.

There is remarkable agreement for $14 < \mu_K < 17$. Brighter than this, the simulations are reaching the limit of the maximum number of stars which can be fit on a frame, and hence they fall short of the theoretical prediction. The flattening at faint surface brightnesses in the simulations is due to only a few objects which are recovered only slightly

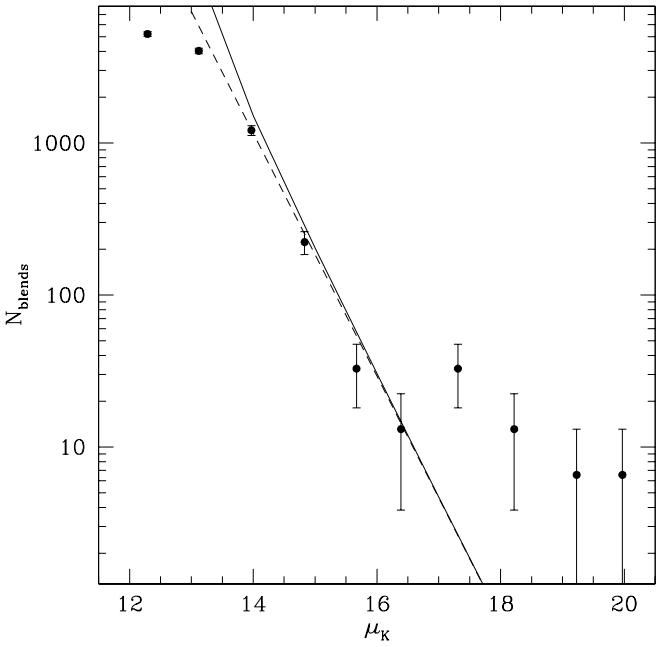


FIG. 13.—Solid line is the number of predicted blends on a NIC2 frame as a function of surface brightness. The dashed line shows the expected number of only two-star RGBT blends. The points are the number of objects measured on each mini-field which are brighter than the brightest input star, normalized to the size of the NIC2 field.

TABLE 5

THEORETICAL POPULATION OF A NIC2 M31 FIELD

		N_j/RES. ELEMENT		N_{blends}/FRAME				
μ_K (mag arcsec^{-2})	L_T ($L_\odot$)	RGBT	RGB	2RGBT	3RGBT	NRGBT	2RGB	M_K
13	5340	0.6	70	7334	4308	17766	...	−7.02
14	2126	0.2	28	1162	272	1517	...	−6.02
15	846	0.09	11	184	17	203	...	−5.02
16	337	0.04	4	29	1	30	...	−4.02
17	134	0.02	2	5	0.07	5	...	−3.02
18	53	0.006	0.7	0.7	0.004	0.7	10561	−2.02
19	21	0.002	0.3	0.1	0.0003	0.1	1674	−1.02
20	8	0.0009	0.1	0.02	0.0000	0.02	265	−0.02
21	3	0.0004	0.04	0.003	0.0000	0.003	42	0.98
22	1	0.0001	0.02	0.0005	0.0000	0.0005	7	1.98

brighter than L_{max}, being blends of an RGBT star with a much fainter star.

Renzini (1998) has given a general (qualitative) criterion for safe stellar photometry in crowded field: "reliable photometry can be obtained only for those stars that are brighter than the average luminosity sampled by each resolution element." The last column in Table 5 gives the total K-band magnitude sampled in M31 by our resolution element. A comparison with Figure 9 shows indeed that the criterion is well confirmed by the detailed simulations. This fairly obvious criterion has been often violated in the past, resulting in a systematic overestimate of the number of stars brighter than the RGB tip in various stellar systems (examples are mentioned in Renzini 1998).

6. IMPLICATIONS FOR FUTURE SPACE-BASED OBSERVATIONS

Even with the improved resolution of new space-based observatories, measurements of distant objects will still be subject to the limitations of image blending. Since the blending threshold is determined by the number of stars present in a resolution element, we can scale our results according to the telescope PSF size and target distance to obtain estimates of the limiting surface brightness for different observations. We need only assume that we have determined the limiting number of stars per resolution element with our NICMOS simulations and that this limit will be similar for other space-based observatories, i.e., assuming similar noise characteristics, PSF structure, and performance as NICMOS. One last consideration (as shown by § 4.2.3) is the effect of the shape and slope of the LF on blending.

As an example, consider measuring Cepheid variables in external galaxies (see Table 6). Accurate determinations of their periods and apparent magnitudes are critical in determining extragalactic distances. For this analysis we assume a mean K-band Cepheid luminosity of $M_K \sim -6$ and ignore other factors such as differences between disk and bulge LFs, metallicities, etc.

Our M31 simulations show that for accurate NICMOS observations of such stars, the K-band surface brightness must be fainter than $\mu_K \sim 15$ mag arcsec^{-2}. This corresponds to a galactocentric distance of 2′ using the r-band Kent (1987) surface brightness measurements, assuming $(r-K) \sim 2.9$ (Terndrup et al. 1994).

What about Cepheid observations in the Virgo cluster? Scaling our M31 K-band NICMOS results to the distance of Virgo (15.9 Mpc), we find that the brightest background where NICMOS can accurately measure Cepheids is $\mu_K = 21.8$. One example of a Virgo spiral is NGC 4548, which in fact has been imaged by the *HST* Key Project team with WFPC2 (Graham et al. 1999). The V-band surface brightness measurements of Benedict (1976), indicate that the Key Project observations occurred in regions with surface brightness $\mu_V \sim 22$. Assuming $(V-K) \sim 1.5$ for a young disk population (Persson et al. 1983), colors typical of SWB III clusters in the LMC in which Cepheids are found, these regions have a K-band surface brightness of $\mu_K \sim 20.5$ mag arcsec^{-2}. Thus the regions imaged with WFPC2 may be too dense for accurate NICMOS photometry because of the infrared instrument's larger PSF.

TABLE 6

CEPHEID BLENDING LIMITS

Target	Distance (Mpc)	Telescope	FWHM (arcsec)	μ_K(lim) (mag arcsec^{-2})
M31	0.7	*HST*-NICMOS	0.15	15.0
M31	0.7	*NGST*	0.04	12.1
Virgo	15.9	*HST*-NICMOS	0.15	21.8
Virgo	15.9	*NGST*	0.04	18.9
Coma.....	102	*NGST*	0.04	22.9

Now consider future observations such as may be performed with a space-based, 8 m–class telescope like the *NGST*. The observations of Cepheids in the Virgo cluster, which are difficult for NICMOS, will now be trivial. Scaling our M31 NICMOS observations to the anticipated *NGST* H-band PSF which will have a FWHM $\sim$0.″04, we find that the limiting K-band surface brightness for accurately measuring Cepheid variables in Virgo is $\mu_K \sim 18.9$. In NGC 4548 this corresponds to a distance of only $\sim$10″ from the nucleus.

More distant targets, such as galaxies in the Coma cluster, will of course be more challenging. To scale from NICMOS observations in M31 to *NGST* in Coma, we again assume that the H-band *NGST* PSF has a FWHM $\sim$0.″04 and that the Coma cluster is at a distance of 102 Mpc. This yields a limiting surface brightness of $\mu_K \sim 22.9$ for accurately measuring Cepheids. Extrapolating the surface brightness analysis of Kent (1987), this limiting surface brightness occurs at $\sim$100′ from the center of M31 (assuming $(r-K) = 2.0$). Thus placing M31 at the distance of the Coma cluster, we find that the inner 0.′7, or $\sim$26 kpc (assuming 1′ = 250 pc in M31), will be inaccessible to even *NGST* for accurately measuring Cepheids in the infrared.

More detailed calculations should of course be undertaken when planning observations, but these scaling arguments give a rough idea of what to expect in terms of the effects of blending.

7. SUMMARY AND CONCLUSIONS

In order to interpret our *HST*-NICMOS observations of globular clusters in M31, we have developed techniques for understanding and quantifying the effects of blending on stellar photometry.

Traditional completeness tests, specifically the injection and recovery of a handful of artificial stars into a preexisting frame, are fine for estimating the photometric completeness. Unfortunately, this type of test does not help in understanding the true brightnesses of stars seen in crowded regions. In very crowded regions, the few added artificial stars are merely added on top of preexisting clumps, with no way of recovering the composition of those clumps.

Our solution is to simulate the entire frame. Our cluster simulations, consisting of hundreds of thousands of stars each (Fig. 1), revealed that stars as bright as we observe near the cluster centers, can be easily created by the blending of many fainter stars (Fig. 6). These simulations can be tailored to nearly any observation and use any desired input LF to best determine the true properties of the observed stellar population, although where blending is severe there are degeneracies in the shape of the input LF.

Using uniformly populated fields of varying surface brightness, we measured the degradation of photometric accuracy as a function of surface brightness. We determined that the photometry of faint stars ($M_K \sim -3$) began to be affected by blending at a surface brightness of $\mu_K \sim 16$ mag

arcsec^{-2}. At a surface brightness of $\mu_K \sim 14$ the photometry of even the brightest stars was affected by blending (Fig. 9). With these criteria, we determined threshold- and critical-blending radii for each cluster, which determine the proximity to each cluster where reliable photometry can be achieved (Table 4).

We show the effects of blending on the measured luminosity function. At low surface brightnesses the measured LF accurately reflects the input LF. However, with increasing surface brightness, the measured LF becomes skewed toward the bright end. At very high surface brightnesses blending dominates the photometry, and the measured LF has a slope opposite in sign to the input LF, actually increasing toward bright stars (Fig. 8).

Blending also affects the measured slope and width of the giant branch. We quantified these effects and determined a relation between the field surface brightness and the artificial increase in the metallicity as derived from the RGB slope (Fig. 11). Results of this modeling will be applied to our M31 cluster and field observations in Paper II.

By scaling our limiting surface brightness imposed by blending, we estimate the limitations blending places on future space-based observations. We show that infrared observations of Cepheid variables with *NGST* will be trivial in the Virgo cluster but may be nearly impossible in the Coma cluster.

Support for this work was provided by NASA through grant number GO-7826 from the Space Telescope Science Institute. R. M. R. acknowledges additional support from NASA contract NAG 5-9431, awarded in connection with the *NGST* ad hoc science working group. We thank Peter Stetson for supplying and helping us with his photometry package, ALLFRAME. Valuable comments from Judy Cohen, Darren DePoy, Paul Martini, and Marcia Rieke were greatly appreciated. We would also like to thank the anonymous referee who provided an extremely thorough and detailed report, which significantly improved this paper.

REFERENCES

Aparicio, A., & Gallart, C. 1995, AJ, 110, 2105
Benedict, G. F. 1976, AJ, 81, 799
Cassisi, S., et al. 2001, in preparation
DePoy, D. L., Terndrup, D. M., Frogel, J. A., Atwood, B., & Blum, R. 1993, AJ, 105, 2121
Frogel, J. A., & Whitford, A. E. 1987, ApJ, 320, 199
Graham, J. A., et al. 1999, ApJ, 516, 626
Hogg, D. W. 2001, AJ, 121, 1207
Kent, S. M. 1987, AJ, 94, 306
Kuchinski, L. E., & Frogel, J. A. 1995, AJ, 110, 2844
Kuchinski, L. E., Frogel, J. A., Terndrup, D. M., & Persson, S. E. 1995, AJ, 109, 1131
MacKenty, J. W., et al. 1997, NICMOS Instrument Handbook, Version 2.0 (Baltimore: STScI)
Persson, S. E., Aaronson, M., Cohen, J. G., Frogel, J. A., & Matthews, K. 1983, ApJ, 266, 105
Renzini, A. 1998, AJ, 115, 2459
Rich, R. M., Mould, J. R., & Graham, J. R. 1993, AJ, 106, 2252
Stephens, A. W., Frogel, J. A., Freedman, W., Gallart, C., Jablonka, P., Ortolani, S., Renzini, A., Rich, R. M., & Davies, R. 2001, AJ, 121, 2597 (Paper II)
Stephens, A. W., Frogel, J. A., Ortolani, S., Davies, R., Jablonka, P., Renzini, A., & Rich, R. M. 2000, AJ, 119, 419
Stetson, P. B. 1987, PASP, 99, 191
———. 1990, PASP, 102, 932
———. 1994, PASP, 106, 250
Stetson, P. B., & Harris, W. E. 1988, AJ, 96, 909
Terndrup, D. M., Davies, R. L., Frogel, J. A., DePoy, D. L., & Wells, L. A. 1994, ApJ, 432, 518
Tiede, G. P., Frogel, J. A., & Terndrup, D. M. 1995, AJ, 110, 2788