7-10-2006

Inverting Color-Magnitude Diagrams to Access Precise Star Cluster Parameters: A Bayesian Approach

Ted von Hippel
University of Texas at Austin, vonhippt@erau.edu

et al.

The Astrophysical Journal, 645:1436–1447, 2006 July 10

INVERTING COLOR-MAGNITUDE DIAGRAMS TO ACCESS PRECISE STAR CLUSTER PARAMETERS: A BAYESIAN APPROACH[1]

Ted von Hippel,[2] William H. Jefferys,[2] James Scott,[3] Nathan Stein,[2] D. E. Winget,[2]
Steven DeGennaro,[2] Albert Dam,[4] and Elizabeth Jeffery[2]
Received 2005 November 7; accepted 2006 March 17

ABSTRACT

We demonstrate a new Bayesian technique to invert color-magnitude diagrams of main-sequence and white dwarf stars to reveal the underlying cluster properties of age, distance, metallicity, and line-of-sight absorption, as well as individual stellar masses. The advantages our technique has over traditional analyses of color-magnitude diagrams are objectivity, precision, and explicit dependence on prior knowledge of cluster parameters. Within the confines of a given set of often-used models of stellar evolution, a single mapping of initial to final masses, and white dwarf cooling, and assuming photometric errors that one could reasonably achieve with the *Hubble Space Telescope*, our technique yields exceptional precision for even modest numbers of cluster stars. For clusters with 50–400 members and one to a few dozen white dwarfs, we find typical internal errors of $\sigma([\mathrm{Fe/H}]) \leq 0.03$ dex, $\sigma(m - M_V) \leq 0.02$ mag, and $\sigma(A_V) \leq 0.01$ mag. We derive cluster white dwarf ages with internal errors of typically only 10% for clusters with only three white dwarfs and almost always $\leq$5% with 10 white dwarfs. These exceptional precisions will allow us to test white dwarf cooling models and standard stellar evolution models through observations of white dwarfs in open and globular clusters.

Subject headings: open clusters and associations: general — stars: evolution — white dwarfs

1. INTRODUCTION

White dwarf cooling theory currently provides the most reliable age for the Galactic disk (Winget et al. 1987; Oswalt et al. 1996; Leggett et al. 1998; Knox et al. 1999), whereas main-sequence stellar evolution provides the most reliable age for the Galactic halo (e.g., Salaris & Weiss 2002; Krauss & Chaboyer 2003). In order to understand the detailed formation sequence of the Galactic components, as well as the local satellite galaxies, these two timescales need to be placed on the same absolute age system. The only current empirical approach available to intercalibrate these two age systems is to derive white dwarf (WD) cooling ages and main-sequence turnoff (MSTO) ages for a number of Galactic star clusters over a wide range of ages and metallicities. Much of the WD age dating work has been necessarily limited to nearby open clusters (Claver 1995; von Hippel et al. 1995; Richer et al. 1998; von Hippel & Gilmore 2000; Claver et al. 2001; von Hippel 2005, hereafter Paper I) that are young or of intermediate age, since old WDs are faint. Hansen et al. (2002) extended WD age studies to one globular cluster (NGC 6121 = M4). They derived a precise WD age, but with large systematic uncertainties due to as yet uncalibrated physical effects in the coolest WDs (Fontaine et al. 2001).

Even though the *Hubble Space Telescope* (*HST*) may be nearing the end of its lifetime, it has made collecting these deep observations of WDs in open and globular clusters possible. At least two more open clusters (NGC 2360 and NGC 2660) and one more globular cluster (NGC 6397) have been observed with *HST* to sufficient depth for the WD technique, and those results will be forthcoming. The large number of 8–10 m telescopes now available will make it possible to observe a few more open clusters to sufficient depth for the WD technique, and the next decade should see 20–30 m telescopes, which will make these studies substantially easier.

While the instrumentation has been improving and there has been steady work on improving WD cooling and traditional main-sequence stellar evolutionary models, there have not been sufficient advances in the statistical machinery available to compare star cluster observations with those models, particularly for WDs. In this paper we present the first phase of our effort to develop this statistical machinery. Specifically, we present a new Bayesian technique that has the ability to objectively incorporate all our prior knowledge, including stellar evolution, star cluster properties, and data quality estimates, while comparing data for each cluster to any available theoretical model. We chose to employ a Bayesian approach precisely because so much is known about stellar evolution and star clusters, and because this approach allows us to test how cluster properties depend on the input models or model ingredients.

The power of the Bayesian approach is impressive, and we show below both the excellent precision that one can obtain in the primary cluster parameters (age, metallicity, distance, and reddening) and the range of related star cluster and stellar evolution problems that can be addressed. The goal of this paper is to present the Bayesian technique and demonstrate its internal precision. In subsequent papers we will derived improved WD and MSTO ages for clusters, with the long-term goal of intercalibrating WD and MSTO ages up to the ages of the oldest globular clusters.

2. BASELINE STELLAR CLUSTER MODEL

We chose a single set of stellar evolution ingredients to build and test the Bayesian approach. We use this model set to test the sensitivity of the derived WD and MSTO ages to the cluster parameters of [Fe/H], A_V, distance, age, number of cluster stars, and assumed photometric error.

For our baseline stellar cluster model we chose a Miller & Scalo (1979) initial mass function (IMF), the main-sequence and giant

[1] WIYN Open Cluster Study XXIII. The WIYN Observatory is a joint facility of the University of Wisconsin–Madison, Indiana University, Yale University, and the National Optical Astronomy Observatory.

[2] Department of Astronomy, University of Texas at Austin, 1 University Station C1400, Austin, TX 78712-0259; ted@astro.as.utexas.edu.

[3] Trinity College, Cambridge CB2 1TQ, UK.

[4] Department of Computer Sciences, University of Texas at Austin, 1 University Station C0500, Austin, TX 78712-0233.

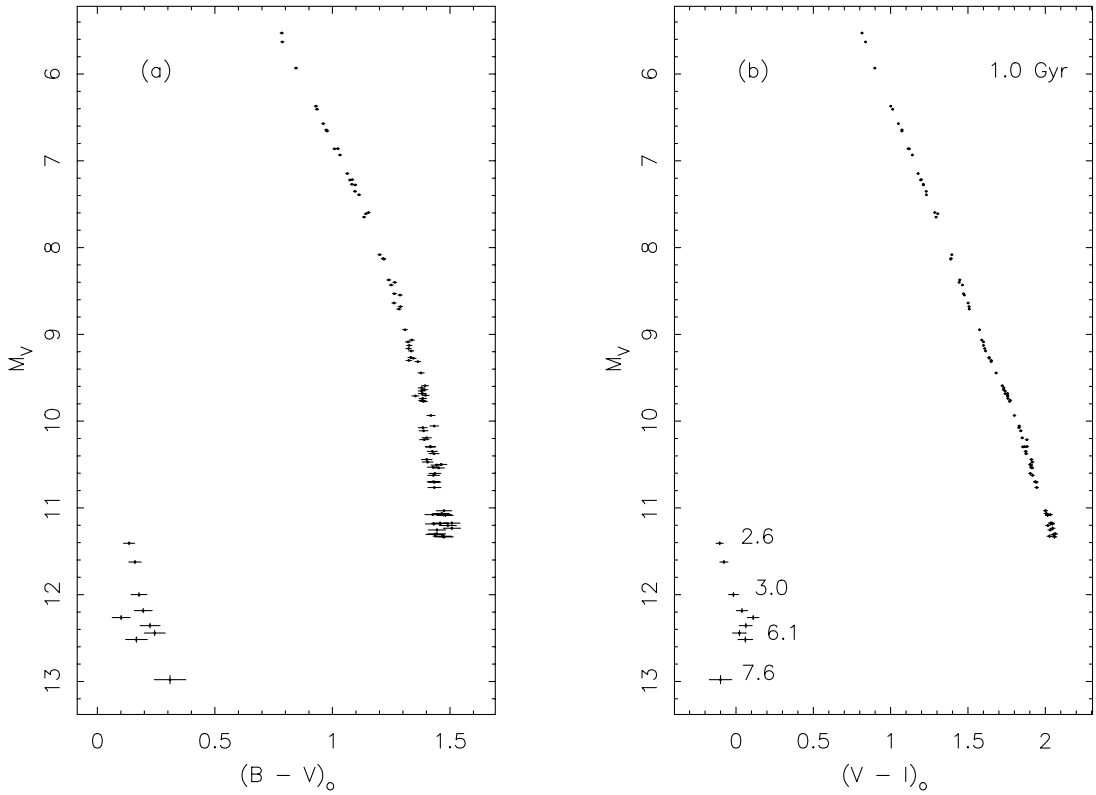

FIG. 1.—(*a*) *BV* and (*b*) *VI* CMDs in the dereddened, absolute magnitude plane for a representative log (age) = 9.0 cluster with [Fe/H] = 0.0, $N = 100$ main-sequence and WD stars, and photometric errors appropriate for $m - M_V = 12.5$ and $A_V = 0$. Photometric errors in the WD region are similar in the *BV* and *VI* CMDs for these simulations, although there is an *x*-axis scale change. Representative zero-age main sequence (ZAMS) masses for the WDs are given in solar units.

branch stellar evolution timescales of Girardi et al. (2000), the initial (main sequence) to final (white dwarf) mass relation of Weidemann (2000), the WD cooling timescales of Wood (1992), and the WD atmosphere colors of Bergeron et al. (1995). Using these ingredients, we simulate star cluster color-magnitude diagrams (CMDs), and, using the Bayesian techniques discussed below, we invert cluster CMDs to recover the probability distribution of the cluster parameters.

When simulating a cluster, each star is randomly drawn from the IMF and, on the basis of a user-specified binary star fraction, is randomly assigned to be a single star or a binary with a companion also randomly drawn from the IMF. Note that although an IMF is required to simulate a cluster, the implied age from either the MSTO technique or the WD technique is insensitive to the IMF. The IMF serves only to increase or decrease the population of stars of interest; for example, MSTO stars or WDs. If there are insufficient stars, particularly if the cluster is young, then the few cluster stars coupled with the IMF can create a statistical uncertainty in locating the MSTO or perhaps even finding WDs. Binaries of nearly any mass ratio have a similar effect. WDs in binaries are generally not recognized, and MSTO stars in such systems are found to be brighter and generally redder than the MSTO and therefore do not help to define the MSTO. For these reasons and for simplicity in this study, we set the binary fraction to 0%, which is substantially lower than the typical value of $\geq$30% for open clusters. For simplicity, we also use only H atmosphere (DA) WDs in our present simulations. While He atmosphere (DB) WDs make up $\leq$10% of field star WD samples (7% in Kleinman et al. 2004), to date no DBs have been found in open clusters (Kalirai et al. 2005). A limitation of our cluster simulations is that stars with masses $\leq 0.25\ M_\odot$ are not included, thus producing an unrealistic lower limit to the main sequence. Since the focus of this study is on stars that can become WDs, this simplification is merely one of presentation.

Other stellar evolution (e.g., Yi et al. 2001; Baraffe et al. 1998; Siess et al. 2000) and WD cooling (e.g., Benvenuto & Althaus 1999; Hansen 1999) models could have been used, and will be added to our code later. For the present purposes, the above-mentioned, often-used models adequately cover parameter space and allow us to build and test the Bayesian machinery.

After producing simulated CMDs, we incorporate realistic photometric errors, assuming reasonable cluster parameters; for example, $m - M_V = 12.5$ and $A_V = 0$–1, and assuming that observations are obtained with the *HST* or a similar imaging instrument able to observe to $V = 27$ with a signal-to-noise ratio of S/N = 15.[5] We use a conservative upper limit to the photometric precision of S/N = 200, although we do not incorporate systematic calibration errors. Our stellar cluster model limits are currently set by the Girardi et al. (2000) and Wood (1992) models, and these limits are 100 Myr–4.5 Gyr and $Z = 0.0004$–0.030 ([Fe/H] ≈ -1.676 to 0.198). This is adequate parameter space for significant age and metallicity exploration and to demonstrate the technique, although we clearly need to push the technique to greater ages.

Our cluster simulations do not include mass segregation or other dynamical processes that are potentially important in open clusters, especially for the lowest mass stars; these typically have little effect on the measured WD mass fraction (von Hippel 1998; see also Hurley & Shara 2003, who find that the WD luminosity function and mass function are insensitive to dynamical effects

[5] From experience, a value of S/N = 15 is required to obtain good morphological rejection of background galaxies at *HST* resolution (von Hippel & Gilmore 2000).

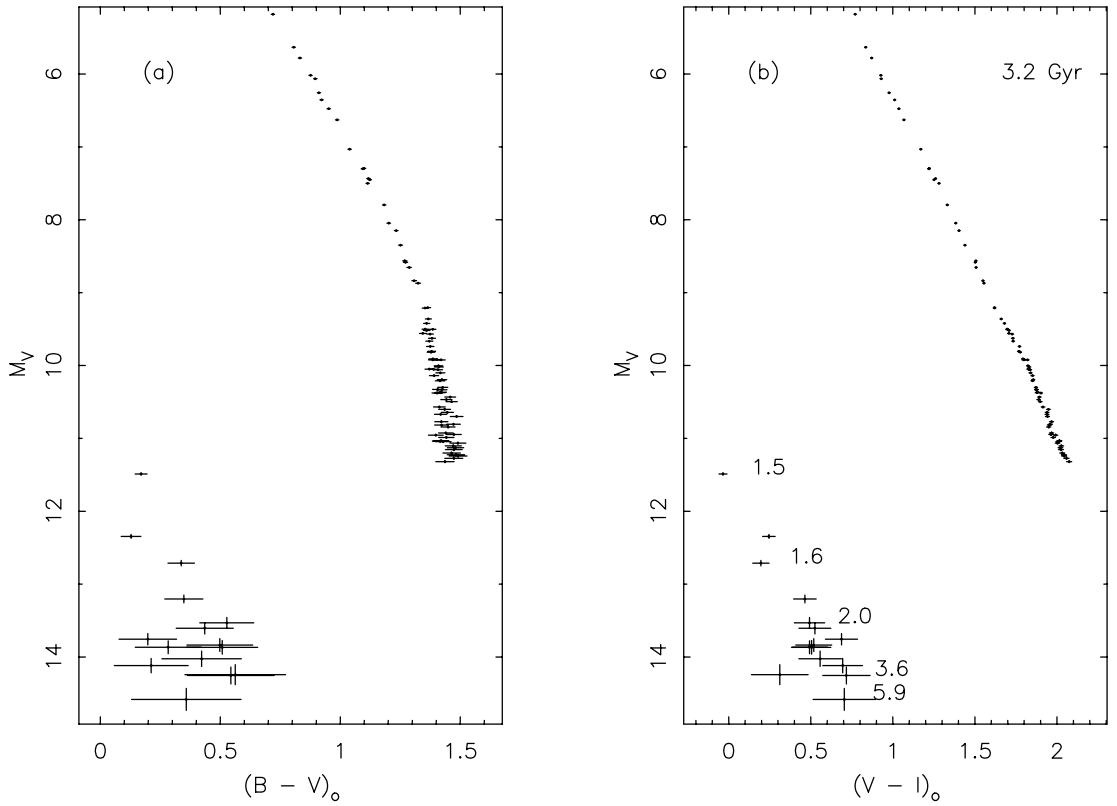

FIG. 2.—Same as Fig. 1, except for log (age) = 9.5. The oldest WDs are now fainter and have larger simulated photometric errors.

at 0.5–1 half-mass radii). Simulated clusters specifically tuned to match real clusters using our stellar cluster model have been presented in Paper I (see its Figs. 4–10). Here we do not attempt to match actual clusters; that is, we do not tune distance, reddening, metallicity, cluster richness, and age, but rather we explore hypothetical clusters that cover the parameter space available to us. The CMDs for two such clusters are presented in Figures 1 and 2 for values of log (age) = 9.0 and 9.5, respectively. The masses of a few WDs from across the cooling sequence are indicated in the right panels of both figures. Between 1 Gyr (Fig. 1) and 3.2 Gyr (Fig. 2), the WD terminus has evolved from $M_V \approx 13$ to $M_V \approx 14.5$, and the simulated photometric errors have increased for the faintest WDs.

In the next section, we outline the Bayesian technique that we will use in forthcoming studies to invert actual CMDs. In verifying the technique, rather than applying our Bayesian code to actual clusters with necessarily unknown parameters, we instead apply our code to simulated clusters. The analyses of simulated data sets test the degree to which an entirely consistent set of stellar models, along with realistic photometric errors, yield the original input parameters. Our Bayesian analyses thus test the internal precision of our technique and its sensitivity to photometric errors, given the many nonlinear aspects of stellar evolution. Since all stellar evolution models are imperfect, this approach provides a measure of internal precision only, not external accuracy. Our goal here is to build a modeling procedure with an internal uncertainty of $\leq 5\%$ in age, which will allow us, when we subsequently analyze real clusters with high-quality data, to test for systematic problems in stellar models and ages of not much more than 5%.

3. BAYESIAN TECHNIQUE

The goal of our Bayesian technique is to use information from the data and from our prior knowledge to obtain posterior distributions on the parameters of our model. Our prior knowledge is encoded in prior distributions on the model parameters. The model parameters include cluster parameters such as age and metallicity and an initial mass for each cluster star. These parameters are the inputs to our stellar cluster model, which we use to derive predicted photometric magnitudes. The likelihood function then compares the predicted magnitudes with the observed (or simulated) data.

Bayes's theorem relates the posterior distribution to the prior distribution and the likelihood function. If $\boldsymbol{M} = (M_1, M_2, \ldots, M_N)$ is a vector of initial masses of all stars in the cluster and $\boldsymbol{\Theta} = (T, [\mathrm{Fe/H}], A_V, m - M_V)$ is a vector of cluster parameters, then we can treat our stellar cluster model as a function $G(\boldsymbol{M}, \boldsymbol{\Theta})$ that maps every reasonable choice of $(\boldsymbol{M}, \boldsymbol{\Theta})$ to a resultant set of photometric magnitudes. To obtain the likelihood, we assume that the errors in our measurements are independently distributed and Gaussian with known variance. Suppose that there are N stars in the cluster and we have observed them through n different filters. Then the observed data form an $n \times N$ matrix $\mathbf{X}$ with typical element x_{ij} representing the magnitude in the ith filter of the jth star. By assumption, each magnitude is normally distributed:

$$x_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2), \tag{1}$$

where μ_{ij} and σ_{ij}^2 are the mean and variance, respectively, of the modeled photometry through filter i of star j. The means and variances also form $n \times N$ matrices, which we call $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The full likelihood is then

$$p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{j=1}^{N} \left(\prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left[\frac{-(x_{ij} - \mu_{ij})^2}{2\sigma_{ij}^2} \right] \right\} \right). \tag{2}$$

The variances $\mathbf{\Sigma}$ come from our knowledge of the precision of our observations. The means $\boldsymbol{\mu}$ are the predicted photometric magnitudes that we obtain from the stellar cluster model:

$$\boldsymbol{\mu} = G(\boldsymbol{M}, \boldsymbol{\Theta}). \tag{3}$$

Thus, the likelihood can be expressed in terms of the variables of our problem and the underlying stellar cluster model:

$$p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{\Sigma}) = p(\mathbf{X}|G, \boldsymbol{M}, \boldsymbol{\Theta}, \mathbf{\Sigma}). \tag{4}$$

Computationally, equation (2) is the most useful form of the likelihood because changing the underlying stellar cluster model leaves equation (2) unchanged. A different stellar cluster model is just a different function, say, H, such that

$$\boldsymbol{\mu} = H(\boldsymbol{M}, \boldsymbol{\Theta}), \tag{5}$$

or even

$$\boldsymbol{\mu} = H(\boldsymbol{M}, \boldsymbol{\Theta}') \tag{6}$$

for a different set of cluster parameters $\boldsymbol{\Theta}'$.

In Bayesian analysis, all model parameters require prior distributions. We have tried to select priors that are consistent with astronomers' knowledge of likely values for the various parameters. To reflect the fact that low-mass stars are much more numerous than high-mass stars, and to be consistent with our stellar cluster model, where we used the Miller & Scalo (1979) IMF, we set the prior distribution on the logarithm of a star's mass to be proportional to the Gaussian distribution:

$$p(\log(M)) \propto \exp\left\{\frac{-[\log(M) + 1.02]^2}{0.917}\right\}, \tag{7}$$

where the constants are from the fit derived by Miller & Scalo and the IMF is bounded at 0.15 and 100 $M_{\odot}$. For metallicity, absorption, and distance modulus, we use Gaussian priors in the common logarithmic versions of these quantities ([Fe/H], A_V, and $m - M_V$). We assume that we have reasonable knowledge of the values and uncertainties of these parameters for a given cluster. This knowledge should come from outside information, not from the color-magnitude data that we intend to analyze. Our prior on T, the base 10 logarithm of the cluster's age, is uniform between $T = 8.0$ and $T = 9.7$ and zero elsewhere. This is a power-law prior on the age with exponent -1, which adequately reflects the observation that younger clusters are more common than older clusters. Note that priors from reliable, previously derived cluster parameters are not required for our Bayesian approach, although they may help. The point is that priors encode any previously determined parameters, where they are available. In some cases, constraining priors [e.g., small σ([Fe/H])] may turn out to be required for precise results; in other cases, such as the ones studied here, constraining priors are unnecessary for precise results.

Given the prior distributions and the likelihood, we obtain the posterior distributions of the parameters from Bayes's theorem, which states that the posterior density $p(\theta|y)$ on model parameters θ given data y is

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{8}$$

where $p(y|\theta)$ is the likelihood and $p(\theta)$ is the prior density on the model parameter θ. The denominator, $p(y)$, is obtained by integrating the numerator over all possible values of θ, such that

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\, d\theta}. \tag{9}$$

In our problem, it is impossible to compute the integral $\int p(y|\theta)p(\theta)\, d\theta$ analytically. Instead, we use the Markov chain Monte Carlo (MCMC) algorithm to approximate the posterior distribution (Casella & George 1992; Chib & Greenberg 1995). The MCMC algorithm allows us to generate a sample from the posterior distribution. We construct a Markov chain such that once it has converged, results of each iteration of the algorithm are distributed approximately according to the posterior distribution, and we regard the history of the chain as a random sample from the posterior. We can thus obtain quantities of interest, such as sample means, without having to analytically compute the normalized posterior distribution.

Our analysis relies on the Metropolis-Hastings algorithm (Chib & Greenberg 1995), which proceeds as follows: Suppose that the current state at iteration t is $\theta^t = \theta$. Propose to move to some new state $\theta^{\star}$. This proposal is generated with density $q(\theta^{\star}|\theta)$. Compute the Metropolis-Hastings factor

$$\alpha = \min\left[\frac{p(\theta^{\star}|y)q(\theta|\theta^{\star})}{p(\theta|y)q(\theta^{\star}|\theta)}, 1\right] \tag{10}$$

and set $\theta^{t+1} = \theta^{\star}$ with probability α. Otherwise, set $\theta^{t+1} = \theta$. Our sample is the parameter sequence $(\theta^n, \theta^{n+1}, \ldots, \theta^N)$, where N is the total number of iterations and n is the number of iterations before the chain converges. We discard the first $n - 1$ iterations, which are referred to as the burn-in. Note the advantage of this method: since $\int p(y|\theta)p(\theta)\, d\theta = \int p(y|\theta^{\star})p(\theta^{\star})\, d\theta^{\star}$,

$$\alpha = \min\left[\frac{p(y|\theta^{\star})p(\theta^{\star})q(\theta|\theta^{\star})}{p(y|\theta)p(\theta)q(\theta^{\star}|\theta)}, 1\right]. \tag{11}$$

We can compute everything in equation (11) without calculating any intractable integrals.

The efficiency of the Metropolis-Hastings algorithm depends heavily on the choice of proposal distribution q. A common choice is a symmetric distribution centered at the current value. This is the "random walk" Metropolis-Hastings sampler. This method has the advantages of simplicity and ease of implementation. However, the sampler can be inefficient if the distribution's width is inappropriate: the sampler might propose excessively small steps and take too long to traverse the parameter space, or it might propose unreasonably large jumps and frequently reject steps. Another option is to choose a proposal distribution that approximates the posterior distribution. This kind of sampler is known as an "independence" sampler, since $q(\theta^{\star}|\theta) = q(\theta^{\star})$, so each proposed value is independent of the current state. The more closely the proposal distribution approximates the target distribution, the higher the acceptance rate and (generally speaking) the more efficient the sampler.

3.1. *MCMC Sampling*

One of the chief problems in designing the MCMC sampler was overcoming the strong correlations between many of the variables. For instance, for a given position on the CMD, an increase in the age of the cluster will require a decrease in the mass of a WD, and vice versa. Since each parameter is sampled individually in sequence, without removing the correlations, the

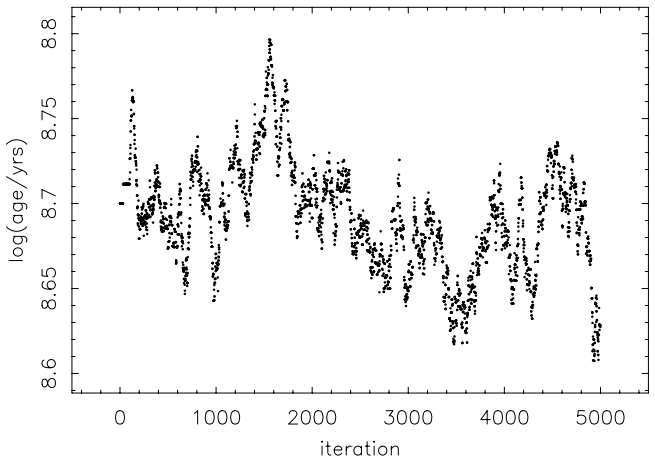

FIG. 3.—Correlated age vs. ZAMS mass sampling for a WD during burn-in.

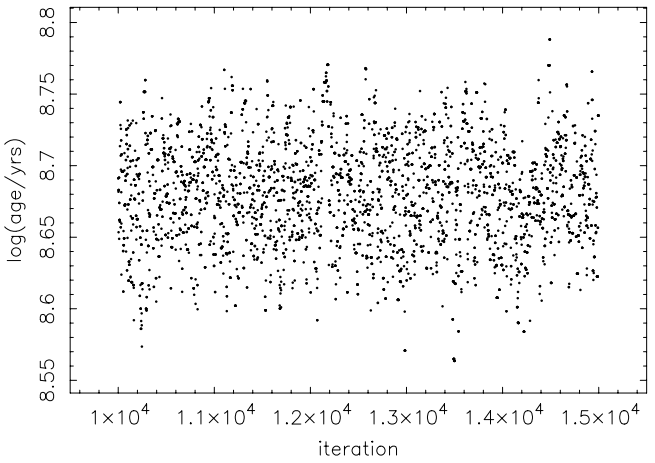

FIG. 4.—Same as Fig. 3, but for the WD after age-mass decorrelation.

sampler can only take small steps in age or mass; if too large a step is taken, the proposed star's photometry will be too far from the observed position and the step will be rejected. In a similar way, metallicity and distance modulus are correlated with each main-sequence star's mass, with each other, and with reddening. While, from a theoretical standpoint, removing these correlations is not required to obtain valid results, the number of iterations needed to be certain that the entire posterior distribution was well sampled would necessitate far more computation time than is practical.

Fortunately, over the ranges that our MCMC algorithm typically samples, these correlations are all nearly linear. To remove the WD age-mass correlation, we introduce a new parameter, U, and a constant, β, defined by

$$M_k = \beta(T_k - \overline{T}) + U_k, \tag{12}$$

where M_k, U_k, and T_k are the mass, decorrelated mass parameter, and logarithm of the cluster age at the kth iteration, respectively, and $\overline{T}$ is the mean logarithmic cluster age. Then, rather than sampling on mass, we sample on U for each star. The MCMC algorithm then computes the mass at each iteration from the above equation. Figures 3 and 4 present the log (age) sampling history before and after the correlation is removed, within the same MCMC run. In Figure 3, age values spanning $\sim$100 iterations are correlated, indicating that little new is learned about the posterior distribution within that correlation length. In Figure 4, the log (age) history is well sampled and each iteration usefully samples the posterior distribution. The new parameter, U, is then decorrelated from distance modulus and metallicity in a similar manner. Finally, the distance modulus and metallicity are decorrelated from one another and then from reddening.

In order to improve the efficiency of our MCMC algorithm, we still need to address several sampling issues. For some parameters, the correlations become nonlinear, often at their extreme values. For other parameters, the correlations consists of two or more separate, nearly linear, pieces with different slopes. For the brightest (youngest) WDs, the correlations between mass and age can be incredibly tight, and further work needs to be done for these objects to more precisely trace these correlations.

The burn-in for our MCMC runs began with a brief (5000 samples) period to approximate the correct values and adjust step sizes. This was followed by two periods of 5000 samples each to calculate the correlation between mass and age for WDs, two more to calculate the correlations between modulus and mass for main-sequence stars and between modulus and absorption, and two more to calculate the metallicity–main-sequence mass and metallicity-absorption correlations. Finally, there is another 5000 sample period to adjust step sizes again. The whole burn-in, except for the initial settling-in period and the first age-mass decorrelation, is then repeated to more precisely determine the correlation factors, for a total burn-in period of 70,000 samples.

4. DEMONSTRATION AND DISCUSSION

For the tests presented here, we placed priors on cluster distance moduli, metallicities, and absorption values. The priors were normal distributions centered on the simulated stellar cluster model parameters for $m - M_V$, [Fe/H], and A_V, with the further requirement that $A_V \geq 0$. The $A_V = 0$ runs are limiting cases, and for these we assumed $\sigma(A_V) = 0$; that is, there was no sampling on A_V. For the $A_V = 0.1$ and 0.3 cases, we assumed $\sigma(A_V) = 0.1$, and for the $A_V = 1$ case, we assumed $\sigma(A_V) = 0.3$. For the other cluster parameters, we assumed $\sigma([\mathrm{Fe/H}]) = 0.3$ dex and $\sigma(m - M_V) = 0.2$. All of these priors represent conservative uncertainties for well-observed, low-reddening clusters. We also placed a prior on the mass distribution for any given star with a form, (see eq. [7]) based on the low-mass IMF.

We simulated B, V, and I photometry for clusters for the range of parameters log (age) = 8.3, 8.7, 9.0, 9.3, and 9.5; [Fe/H] = -1.0, -0.15, 0.0, and 0.15; N (number of cluster stars fainter than the MSTO, including WDs) equal to 50, 100, 200, and 400; $m - M_V = 12.5$; and $A_V = 0$, 0.1, 0.3, and 1. Since our goal is to test the age sensitivity of the WDs, we removed all MSTO, subgiant, and giant branch stars from each simulation, so that what remains are WDs and essentially unevolved main-sequence stars.

While we astronomers are most comfortable studying star clusters in the CMD, our Bayesian technique does not use the CMD, with its correlated errors between the x-axis and y-axis, but rather uses an n-dimensional space, where n is the number of filters available and the units are magnitudes. In the numerical experiments we present here, $n = 3$, as we use B, V, and I photometry. The MCMC routine sees the CMD of Figure 1, for example, in a form more akin to that shown in Figure 5, although offset by the simulated distance modulus. For presentation purposes, we reduced three dimensions to two by plotting either the B or I absolute magnitudes on the horizontal axis. One disadvantage of this plot is the large dynamic range in both axes. Still, the main CMD features can be discerned; for example, WDs are clearly visible in the faintest corner of the plot in Figure 5. Reddening vectors for $A_V = 1.0$ are also shown, as is the effect of increasing distance modulus by 1.0 mag. Both distance and

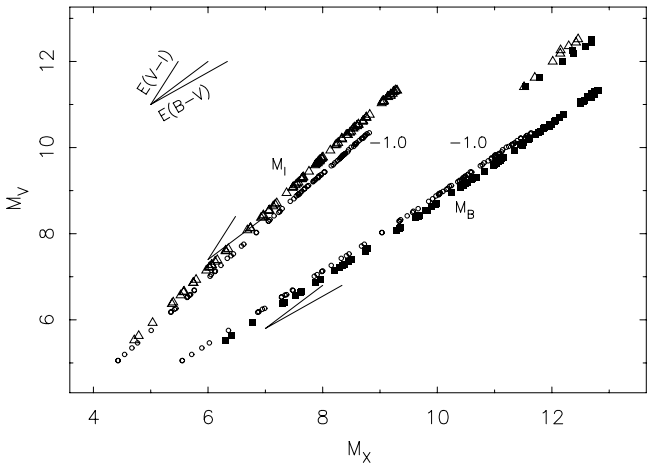

FIG. 5.—Bayesian code version of the CMDs in Fig. 1. Filled squares represent the *BV* plane and open triangles represent the *VI* plane for solar metallicity stars. Cluster WDs are in the upper right, with the *VI* sequence slightly above the *BV* sequence. The effects of metallicity are shown by the open circles, which are the *BV* and *VI* main sequences for $[\mathrm{Fe/H}] = -1$. The offsets of distance and reddening are removed from this plot for presentation purposes. Reddening vectors for $A_V = 1$ for both planes are given in the upper left, as is a vector showing the effect of increasing the cluster's distance modulus by 1 mag. The reddening vectors and distance offset vector are replotted near the main sequences to facilitate comparison.

the reddening vectors are nearly parallel to the main sequence, especially the *BV* main sequence. Decreasing metallicity from $[\mathrm{Fe/H}] = 0.0$ to -0.1 moves the main sequences in almost the opposite directions as the reddening vector. While there are some morphological features in this diagram, and while the various distance, reddening, and metallicity vectors are not absolutely parallel and therefore not entirely degenerate, this diagram suppresses subtleties that primarily affect stellar color.

Although we can simulate clusters younger than log (age) = 8.3, the MCMC technique requires sampling an age range, and for younger clusters this would often hit our (current) lower age limit of log (age) = 8.0. For the A_V cases, three simulated clusters were run for each unique set of cluster parameters. Any two clusters with identical parameters will yield different CMDs as both the IMF and the simulated photometric error distribution are sampled anew. After creating a cluster, we pass it to the MCMC routine with estimates of the mass of each star and estimates of

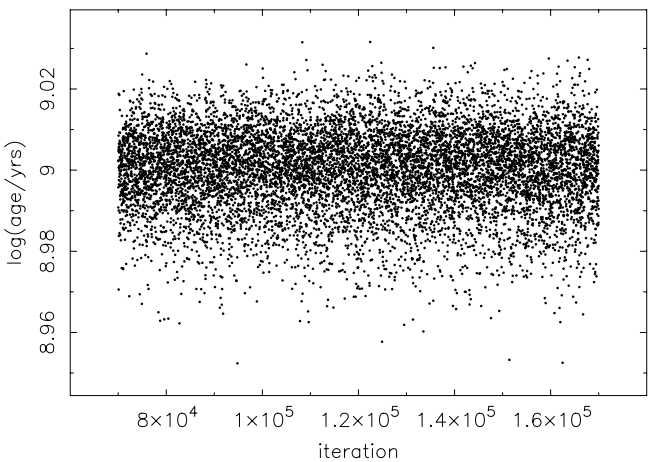

FIG. 6.—Typical history plot of cluster log (age), shown here for the 1 Gyr cluster of Fig. 1. In this case, the sampling was excellent. For clarity, only every 100th point is plotted, and only after the initial 70,000 iteration burn-in period.

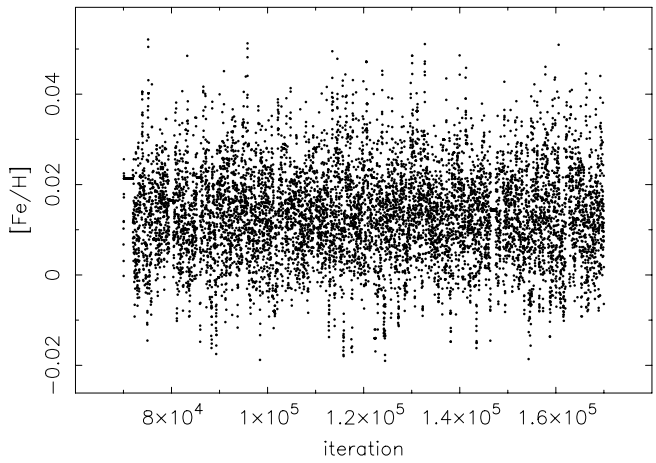

FIG. 7.—Similar to Fig. 6, but for cluster [Fe/H]. This particular MCMC run had a very mild correlation in [Fe/H].

the cluster parameters as starting points. (Our experiments show that as long as the MCMC algorithm converges, the results do not depend on the starting points. Starting points within a factor of ~2 in age or metallicity, for instance, are adequate for convergence.) Because the MCMC sampling is still often correlated, we sample for 10^6 iterations, reading out every 10th value for each stellar mass and for the cluster parameters. Many of our MCMC runs have a correlation length of ≤ 10, and this produces uncorrelated parameter values. For those cases that still remain correlated, and guided by the rule of thumb that one typically wants 10^4 uncorrelated iterations in order to adequately sample the posterior distribution, we find that 10^6 iterations works well for most of our simulated clusters.

Figure 6 presents a well-sampled, typical history plot of cluster age for the cluster of Figure 1. Figures 7 and 8 present the companion history plots for cluster [Fe/H] and $m - M_V$. There is a small amount of sticking in the sampling of these two variables at the beginning of the sequence, just after burn-in, and again near iteration 1.47×10^5, but otherwise these history plots are well sampled. Since there is no A_V sampling in the $A_V = 0$ case of Figure 1, Figure 9 presents the A_V history plot for a cluster with the same parameters, except with an input value of $A_V = 1$. Histograms of these four types of history plots (Fig. 10) are the estimates of the posterior probability distributions. In Figure 10 we present also the $A_V = 0.3$ and 1 cases. Figure 10 shows that the posterior distributions of log (age) and the other cluster

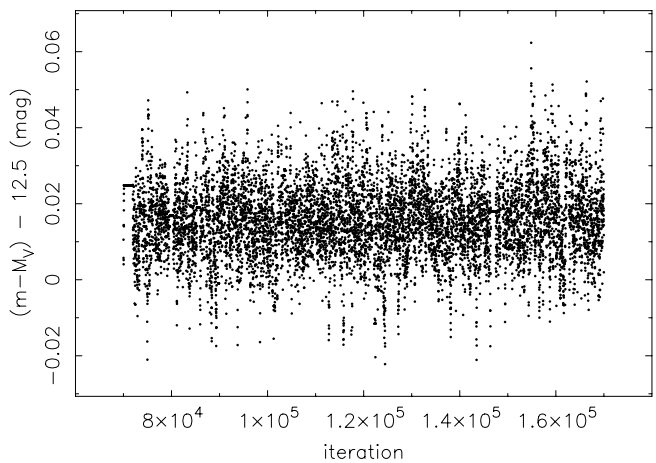

FIG. 8.—Similar to Fig. 6, but for $m - M_V$.

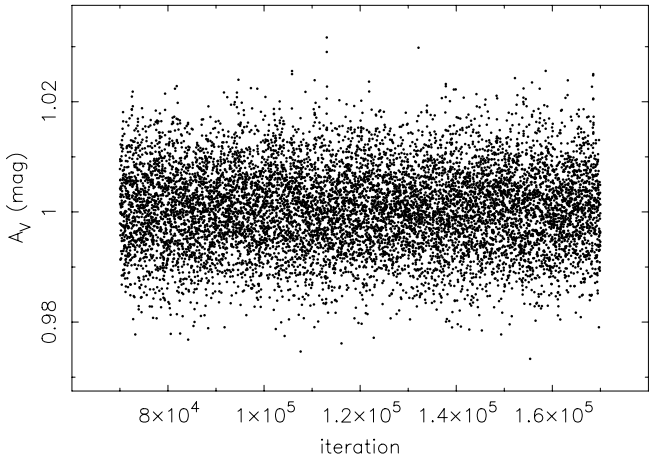

FIG. 9.—Similar to Fig. 6, but for A_V in a cluster with $A_V = 1$.

parameters are close to normal, and furthermore that changing the absorption causes no strong bias in the results (more on absorption and bias below). From these posterior probabilities we can calculate statistics of interest (e.g., mean, median, σ, percentiles), and these are presented below. Figure 11 presents the posterior probabilities of mass for four stars from the cluster simulation of Figure 1. The top left panel shows the mass posterior for a high-mass WD, the top right panel shows that for a lower mass WD, the bottom left panel shows that for a main-sequence star not far below the turnoff, and the bottom right panel presents that for a low-mass main-sequence star. In all cases the mass distribution is nearly centered on the input mass, the mass value before its photometry was subject to random error, except in the case of the lowest mass star, where it differs by only 0.002 $M_\odot$. For the main-sequence stars, the mass distribution is particularly narrow, showing that within the assumption of a specific model, precise photometry yields precise masses. The WD mass distributions are slightly broader because we plot the zero-age main sequence (ZAMS) masses for these stars and a wider range of initial main-sequence masses is converted into a narrow range of WD masses via the initial-final mass relation (Weidemann 2000).

In Figure 12, we check the differences between the mean log (age) values of the distributions and the input log (age) values compared to the standard deviations of the posterior log (age) distributions (σ). We are essentially asking what the deviation of each result is in units of its standard deviation. For the $A_V = 0$ case, the distribution of errors is very similar to the overplotted normal distribution. Formally, the error distribution is also close to normal, with an average of 0.037, a median of 0.055, a standard deviation of 0.985, and a skew of −0.191. This comparison is a sanity check on the self-consistency of our implementation of the Bayesian technique and whether the standard deviation statistic adequately captures the shapes of the posterior distributions. For the $A_V = 0.1$, 0.3, and 1 cases, our MCMC approach tends to be biased high in age by 0.56 σ, 0.26 σ, and 0.12 σ, respectively, and the higher absorption cases tend to have a pronounced nonnormal distribution. These distributions are virtually identical if we plot median values instead of the means of the posterior distributions. While we are still trying to understand some of the subtleties of the higher A_V cases, these offsets, which correspond to 2.6%, 1.2%, and 0.6% systematic errors in age, are small enough that we set aside their

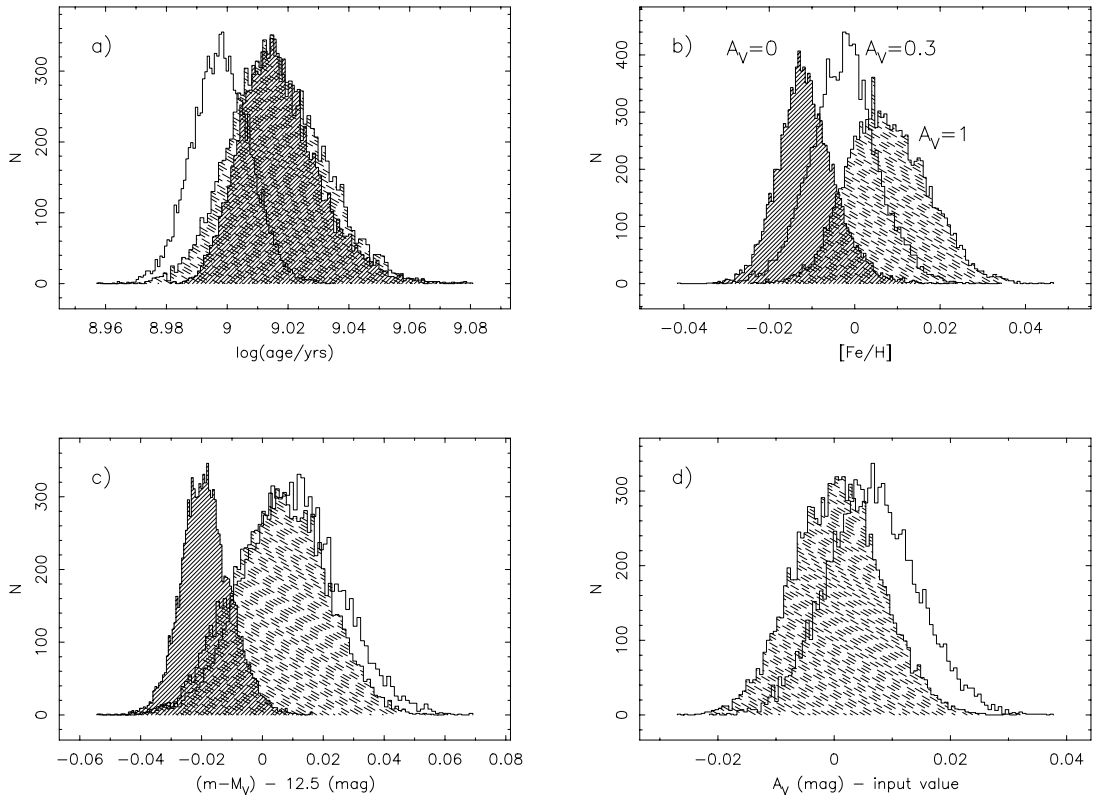

FIG. 10.—Histograms of the MCMC history plots, for (*a*) the log (age) values of Fig. 6, (*b*) the [Fe/H] values of Fig. 7, (*c*) the $m - M_V$ values of Fig. 8, and (*d*) the A_V values of Fig. 9. For the first three panels, the posterior distributions for $A_V = 0$, 0.3, and 1 are plotted, and for the final panel the latter two distributions are plotted, since there was no sampling on A_V for the $A_V = 0$ case. Every 10th iteration, which is the frequency with which these MCMC runs were read out, is incorporated into the histograms.

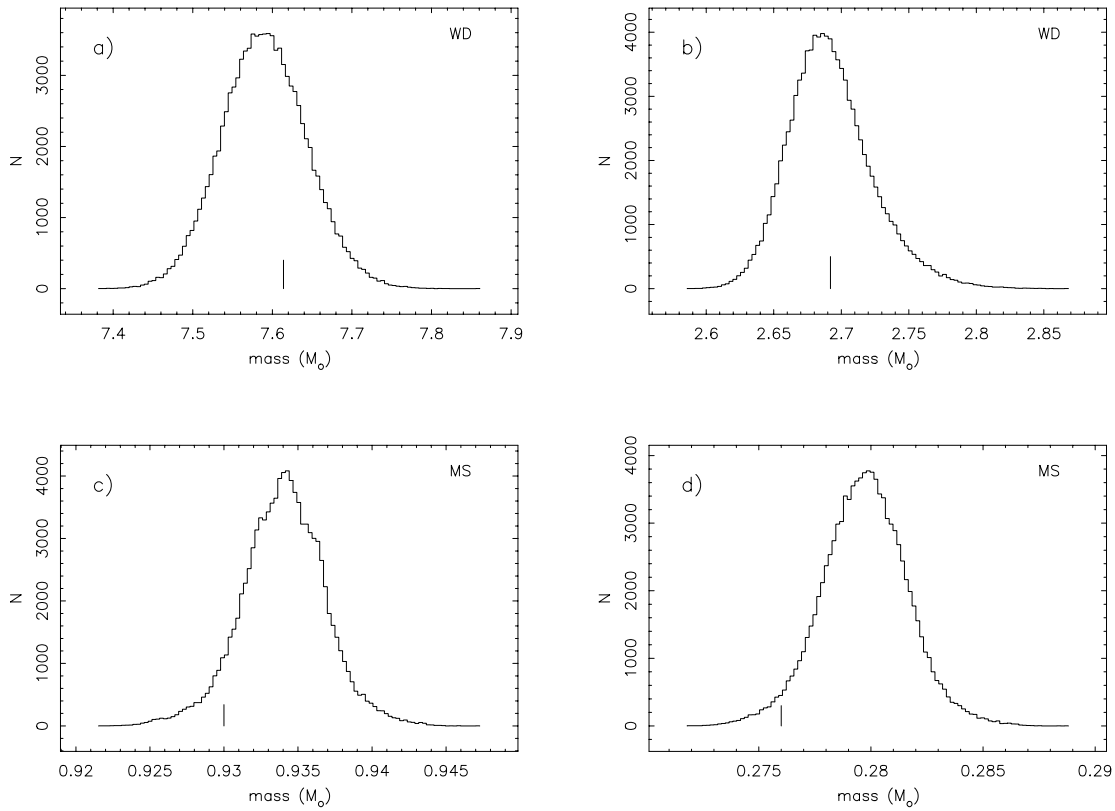

FIG. 11.—Posterior probabilities of ZAMS masses for four stars from the cluster presented in Fig. 1. Panel (*a*) shows the mass posterior for a high-mass WD, panel (*b*) shows that for a lower mass WD, panel (*c*) shows that for a main-sequence star not far below the MSTO, and panel (*d*) shows that for a low-mass main-sequence star. In all cases, the means of the mass distributions are similar to the input masses, labeled with small vertical marks at the bottom of each panel.

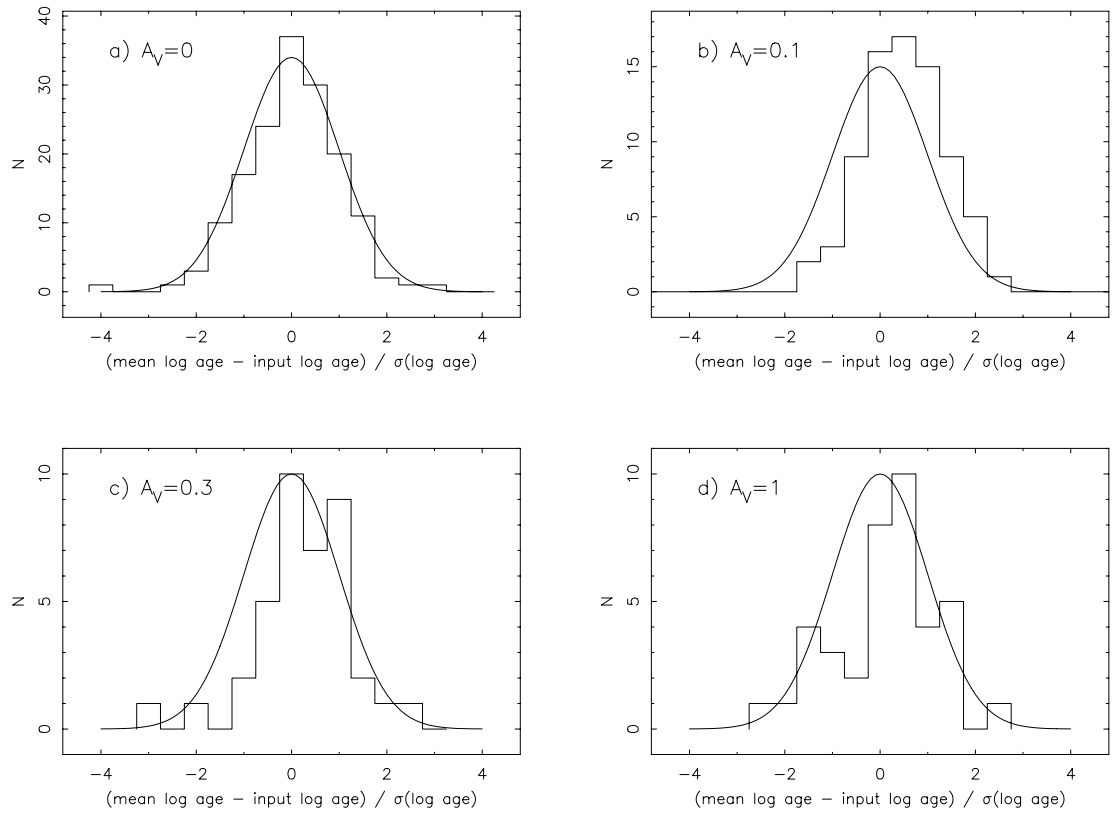

FIG. 12.—Differences between the mean ages of the posterior distributions and the input ages, normalized by the standard deviations (σ) of each posterior age distribution, for the (*a*) $A_V = 0$, (*b*) $A_V = 0.1$, (*c*) $A_V = 0.3$, and (*d*) $A_V = 1$ cases. The distribution of errors is very similar to the overplotted normal distribution for the $A_V = 0$ case and becomes subtly biased to 0.12–0.56 σ for higher values of A_V.

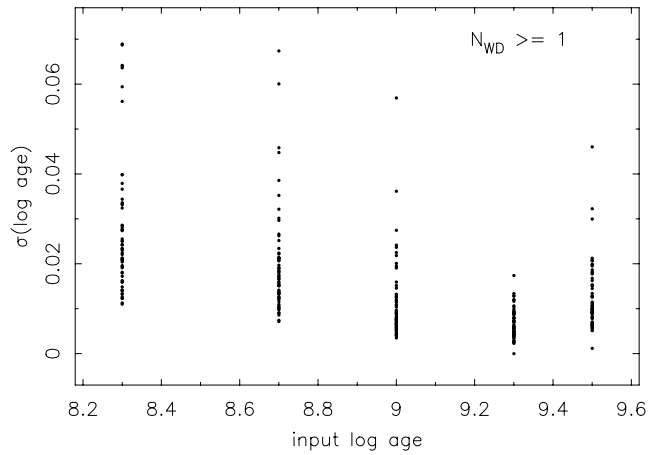

FIG. 13.—Derived standard deviations for each model with one or more WDs vs. the actual (input) ages of the clusters. The standard deviations are always small, typically $\leq$0.04, corresponding to relative errors of typically $\leq$10% for all ages tested. One result (among 319 runs) with $\sigma(\log(\text{age})) = 0.105$ at log (age) = 8.7 and $N_{\rm WD} = 1$ is not plotted for presentation purposes.

resolution for now. Figure 12 demonstrates, particularly for low absorption values, that the standard deviations in the posterior distributions are accurate assessments of the uncertainties due to photometric errors or any effects due to small number statistics, such as having very few WDs in young or sparse clusters.

In Figure 13, we present the standard deviation log (age) uncertainty for each model with one or more WDs versus the input log (age) value of the cluster. The standard deviations are always small, typically $\leq$0.04 dex, corresponding to relative errors typically $\leq$10%, for all ages tested. These errors do not depend significantly on the cluster age. In fact, the apparent slight dependence on age seen in Figure 13 is a combination of two other effects: there are fewer WDs in the youngest clusters, and the coolest WDs in the oldest clusters are fainter and therefore have higher photometric errors. Figure 14, which plots the same standard deviation uncertainties, now versus log ($N_{\rm WD}$), shows the most important factor in this technique: the number of WDs per cluster. Although every WD contains age information, the quality of that information is not the same for all WDs. Photometry for the coolest WDs in any cluster provides the most information (see below), yet photometric precision drops

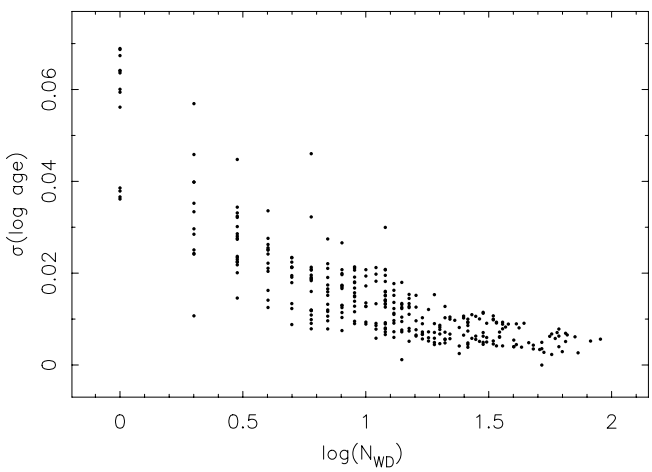

FIG. 14.—Standard deviation uncertainties vs. log ($N_{\rm WD}$) for the data presented in Fig. 13. The precision of the age fit improves approximately as the logarithm of the number of WDs, which is an important factor under the observer's control.

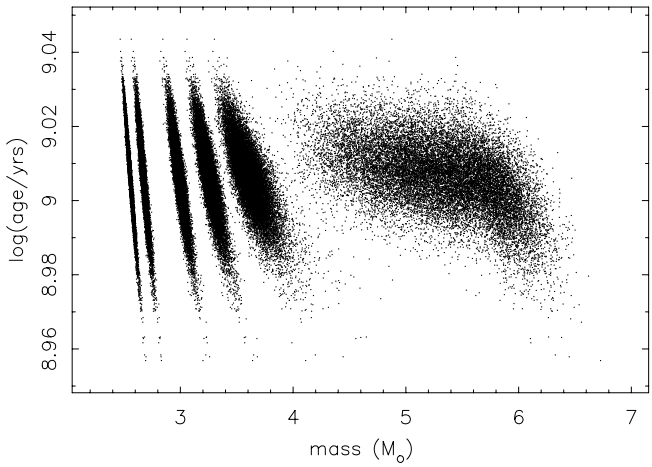

FIG. 15.—WD ZAMS masses vs. cluster age for six of the nine WDs for the cluster in Fig. 1. The other three WDs are not plotted for clarity.

with decreasing WD luminosity. Therefore the age precision does not improve as the square root of the number of WDs, but somewhat more slowly. Nonetheless, even with 10 WDs the statistical (internal) error is almost always $\leq$0.02 dex, or $\leq$5%. Even with three WDs, 10% precision is usually achieved.

Since the precision is so high via this technique, it is worth taking a small detour into the details of the age information locked up in each WD. Figure 15 presents the relationship between possible masses and possible ages for six of the nine WDs of Figure 1 [log (age) = 9.0]. Two WDs are not presented because they are so close in mass to other WDs that they crowd the figure without presenting any new information. The faintest WD of Figure 1 has been dropped also, since its ZAMS mass (7.6 $M_\odot$) is beyond the 7.0 $M_\odot$ limit of the Weidemann (2000) initial-final mass relation and its WD mass (1.131 $M_\odot$) is beyond the Wood (1992) 1.0 $M_\odot$ cooling WD model limit. Although this extrapolated WD has properties that are internally consistent enough for the MCMC runs, they are not pedagogically helpful in understanding the precision in the WD technique. Figure 15 shows that the age-mass relationship for the hotter, brighter WDs is highly correlated, which is the cause of the numerical difficulties with the correlated sampling mentioned above. Is there useful age information in these hot, rapidly cooling WDs, information that is unexploited in the traditional approach, which uses the coolest WDs to derive a cluster's age?

Figure 16 presents the age sensitivity of this technique[6] by presenting the allowed age-mass relationship for each of the WDs of Figure 15, *if each were the only WD in the otherwise identical cluster*. Running the same cluster with only one WD yields the age constraints from an individual WD, while still relying on the cluster main sequence to constrain the combination of metallicity, distance, and reddening. Now it is clear that the higher mass WDs provide the tightest age constraints, eliminating a significant age range allowed by the lowest mass, hottest WD, for example. Yet, in this case even the second hottest WD contains significant age information. Our technique can be pushed to the point at which ages can be derived for

[6] Note that the slight changes of slope in the mass-age relationship for the lowest mass WD is a numerical artifact caused by our use of linear interpolation among the Girardi et al. (2000) models. Besides being a small effect and outside of the actual fit presented in Fig. 15, choppiness due to linear interpolation serves to occasionally slightly decrease the precision of our technique. Higher order interpolations have not yet been necessary and would nearly double the run time of our MCMC code.

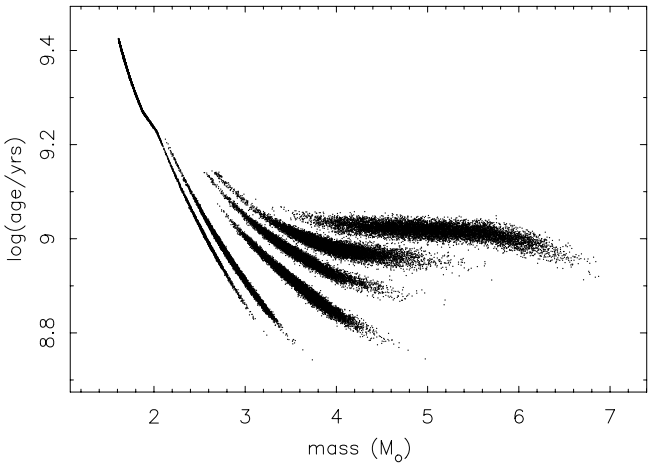

FIG. 16.—WD ZAMS masses vs. cluster age for six modified versions of the cluster presented in Fig. 1. In order to explore the mass-age correlations and see which WDs provide the greatest age constraints, nine clusters, each with only one WD from the original nine cluster WDs, were created. Again, only six of these mass-age relations are plotted for clarity. The lowest mass WDs have the tightest mass-age correlations, which creates greater MCMC sampling challenges. The higher mass WDs provide tighter age constraints. The kinks in the lowest mass relationship occur at boundaries of the main-sequence (Girardi et al. 2000) tracks and are numerical artifacts.

clusters without observing the coolest WDs, and a companion paper will explore the sensitivity of that approach (Jeffery et al. 2006). The higher mass WDs have a much flatter slope in the mass-age diagram, since large changes in ZAMS mass do not appreciably change the contribution of precursor timescales, nor do they evolve at a much different rate than lower mass WDs, at least not in this age range.

Figure 16 still begs the question: Why is there any age sensitivity when there is only one cluster WD? The short answer to this is that the WD region of the CMD is not highly degenerate. Although it may be possible to make a highly degenerate CMD with a combination of few cluster stars, high and uncertain reddening, uncertain metallicity, etc., generally this is not the case. There are also other constraints on the WD properties. A WD cannot have a mass higher than the upper limit for creating WDs (8 $M_\odot$ in all our simulations, most likely 7–9 $M_\odot$), and a WD cannot have a mass so low that stars with lower initial masses are still present on the main sequence. Changes in mass move a WD in the CMD along essentially the same vector as changes in age for hot WDs, where precursor ages are important, and in a perpendicular direction to age for cooler WDs, where precursor ages are unimportant. Figure 17 attempts to make this clear by plotting a small portion of the CMD of Figure 1 around the simulated WDs. The simulated WDs are presented as error bars, whereas the input values, before photometric scatter was added, are presented as filled circles. Here all cluster WDs, except the highest mass (7.6 $M_\odot$) simulated WD, are plotted. The small plus signs connected by lines show the effect of holding mass constant while changing log (age) by ± 0.01 dex, or, in the case of the two highest mass WDs, by changing log (age) by ± 0.02 dex. Open squares show the effect of keeping log (age) $= 9.0$ while adjusting the ZAMS masses by $\pm 2\%$, or, for the two highest mass WDs, by $\pm 5\%$. WD isochrones for log (age) $= 8.9$, 8.95, 9.0, 9.05, and 9.1 are overplotted. Ultimately, the Bayesian technique is so sensitive because minor changes in WD mass or cluster age of just a few percent move the expected location of

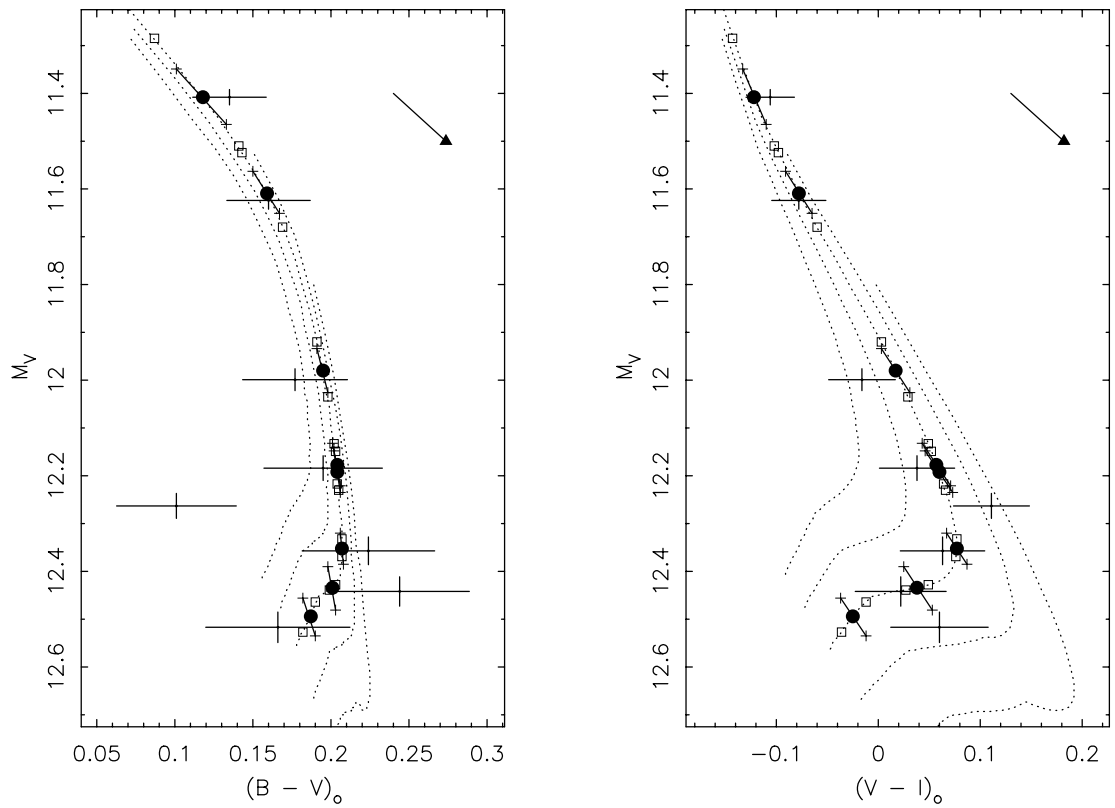

FIG. 17.—WD regions of the CMDs in Fig. 1. The input WDs are plotted as filled circles, and the scattered photometry data are plotted as 1 σ error bars. The highest mass WD is not plotted (see text). The plus signs connected by lines show the effect of changing log (age) by ± 0.01 dex, or, in the case of the two highest mass WDs, by ± 0.02 dex. Open squares show the effect of changing ZAMS masses by $\pm 2\%$, or, for the two highest mass WDs, by $\pm 5\%$. WD isochrones for log (age) $= 8.9$, 8.95, 9.0, 9.05, and 9.1 are overplotted. The reddening vectors for $A_V = 0.1$ are also shown.

any WD significantly in the CMD. Also, some types of photometric error, such as the $\sim 2\sigma$ color error in $B - V$ of one of the WDs in the middle of the cooling sequence, cannot be matched by any realistic adjustment of cluster or stellar parameters, and thus this photometric error does not drive the fit for this WD. In Figures 15 and 16, for example, this WD is the object plotted third from the right. Its ZAMS mass (3.285 $M_{\odot}$) and age [log (age) = 9.0] sit right in the center of the sampled values, so this color error had no meaningful effect. Errors in color could cause one to mistake a WD for a field star, however. The solution to this problem in real clusters with possibly contaminating field stars is better photometry or classification-level spectroscopy to confirm the WD.

Besides deriving precise ages, the Bayesian technique also can derive precise values for the cluster parameters of metallicity, $m - M_V$, and A_V. In all these cases the standard deviations in the posterior distributions are small, typically $\leq$0.03 dex, $\leq$0.02 mag, and $\leq$0.01 mag, respectively. All of these posterior uncertainties are an order of magnitude smaller than the width of the prior distributions (0.3 dex, 0.2 mag, and 0.1–0.3 mag, respectively), demonstrating that high-quality priors in these parameters are not generally needed, at least for low to moderate, single-valued absorption, and that the results are insensitive to the exact assumed starting values for these parameters. The cluster photometry contains a wealth of information, and the Bayesian technique, along with an assumed model set, brings this out to high precision. For clusters with 10 WDs, the age precision is typically better than 5%, easily meeting our needs for a precise statistical tool.

5. BAYES MEETS STAR CLUSTERS: OTHER USES

In our work to date, we have focused on cluster ages, particularly via the WD technique. Age via the MSTO technique will be next. Our MCMC code also derives values for the other major cluster parameters: metallicity, distance modulus, and line-of-sight absorption, along with the individual stellar property of mass. For our own purposes, we intend to upgrade our MCMC code to include binaries drawn from realistic mass ratio distributions, field stars, DB (He atmosphere) WDs, and a wider range of standard stellar and WD evolution models, including ages up to globular cluster values.

Here are a few further example uses for our Bayesian technique:

1. In our effort to improve WD and MSTO ages, we will systematically study main-sequence and WD model parameters that affect ages, such as core convection prescriptions, nonstandard elemental abundances, and diffusion in main-sequence stars, as well as surface convection prescriptions and C/O phase separation in cool WDs.

2. We intend to study the sensitivity in the implied underlying parameters of simulated and actual clusters to the initial-final mass relation, as well as the upper mass limit for creating WDs.

3. Our code derives mass posterior distributions for every object in the cluster. These mass estimates would be a good starting point for IMF studies, particularly since one can adjust the priors on mass to reflect different assumed IMFs and one can incorporate as input any stellar evolution model. By adjusting the prior on the IMF, one could see how many cluster stars are required before the resulting IMF is no longer sensitive to the prior.

4. Once we add binaries to the MCMC code, we intend to study cluster binaries, their masses, and mass ratios. This would be a step forward from the typical approach of estimating cluster binary contributions by visually studying the distribution of stars above the single-star main sequence. Additional information, such as the probability of cluster membership from proper motion or radial velocity studies, could also be incorporated in the binary studies.

We expect to make our code publicly available within a year, after it passes out of its development stage.

6. CONCLUSIONS

We have demonstrated a new Bayesian technique to invert color-magnitude diagrams to reveal the underlying cluster properties of age, distance, metallicity, and line-of-sight absorption, as well as individual stellar masses. We do not fit cluster fiducial sequences, nor do we create plots with many combinations of cluster parameters and then try to derive the best parameters via chi-by-eye. The Bayesian technique delivers not just parameters and error estimates, but entire posterior distributions. Posterior distributions for the parameters of interest are particularly valuable when they may be nonnormal, as may occur with all the coupled, nonlinear aspects of stellar evolution. Despite the potential for complex error distributions, we find posterior age distributions that are close to normal in log (age). Some other distributions, for example, some mass distributions, are clearly nonnormal.

Within the confines of a given set of often-used models of stellar evolution, the initial-final mass relation, and WD cooling, and assuming photometric errors that one could reasonably achieve with *HST*, we find that our technique yields exceptional precision for even modest numbers of cluster stars. For clusters with 50–400 members and one to a few dozen WDs, we find typical internal errors of $\sigma([\mathrm{Fe/H}]) \leq 0.03$ dex, $\sigma(m - M_V) \leq 0.02$ mag, and $\sigma(A_V) \leq 0.01$ mag. The parameter we are most concerned with, cluster WD age, has an internal error of typically only 0.04 dex (10%) for clusters with only three WDs and almost always $\leq$0.02 dex ($\leq$5%) with 10 WDs. All of these results have posterior distributions that are an order of magnitude narrower than the priors we applied and therefore represent the actual information in cluster CMDs. Cluster photometry clearly contains a wealth of information, much of it coupled in a nonlinear fashion, and the Bayesian technique, along with an assumed model set, brings this out to high precision.

We thank the referee, Gordon Drukier, for helpful suggestions that improved our manuscript. This material is based on work supported by the National Aeronautics and Space Administration under grant NAG5-13070 issued through the Office of Space Science. We also gratefully acknowledge a grant to J. S. from the NSF-funded Vertical Integration of Research and Education (VIGRE) program awarded by the Department of Mathematics of the University of Texas.

REFERENCES

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403
Benvenuto, O. G., & Althaus, L. G. 1999, MNRAS, 303, 30
Bergeron, P., Wesemael, F., & Beauchamp, A. 1995, PASP, 107, 1047
Casella, G., & George, E. I. 1992, Am. Statistician, 46, 167
Chib, S., & Greenberg, E. 1995, Am. Statistician, 49, 327
Claver, C. F. 1995, Ph.D. thesis, Univ. Texas, Austin
Claver, C. F., Liebert, J., Bergeron, P., & Koester, D. 2001, ApJ, 563, 987
Fontaine, G., Brassard, P., & Bergeron, P. 2001, PASP, 113, 409
Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371
Hansen, B. M. S. 1999, ApJ, 520, 680

Hansen, B. M. S., et al. 2002, ApJ, 574, L155
Hurley, J. R., & Shara, M. M. 2003, ApJ, 589, 179
Jeffery, E. J., von Hippel, T., Jefferys, W. H., Winget, D. E., Stein, N., & DeGennaro, S. 2006, ApJ, submitted
Kalirai, J. S., Richer, H. B., Hansen, B. M. S., Reitzel, D., & Rich, R. M. 2005, ApJ, 618, L129
Kleinman, S. J., et al. 2004, ApJ, 607, 426
Knox, R. A., Hawkins, M. R. S., & Hambly, N. C. 1999, MNRAS, 306, 736
Krauss, L. M., & Chaboyer, B. 2003, Science, 299, 65
Leggett, S. K., Ruiz, M. T., & Bergeron, P. 1998, ApJ, 497, 294
Miller, G. E., & Scalo, J. M. 1979, ApJS, 41, 513
Oswalt, T. D., Smith, J. A., Wood, M. A., & Hintzen, P. 1996, Nature, 382, 692
Richer, H. B., Fahlman, G. G., Rosvick, J., & Ibata, R. 1998, ApJ, 504, L91
Salaris, M., & Weiss, A. 2002, A&A, 388, 492
Siess, L., Dufour, E., & Forestini, M. 2000, A&A, 358, 593
von Hippel, T. 1998, AJ, 115, 1536
———. 2005, ApJ, 622, 565 (Paper I)
von Hippel, T., & Gilmore, G. 2000, AJ, 120, 1384
von Hippel, T., Gilmore, G., & Jones, D. H. P. 1995, MNRAS, 273, L39
Weidemann, V. 2000, A&A, 363, 647
Winget, D. E., Hansen, C. J., Liebert, J., van Horn, H. M., Fontaine, G., Nather, R. E., Kepler, S. O., & Lamb, D. Q. 1987, ApJ, 315, L77
Wood, M. A. 1992, ApJ, 386, 539
Yi, S., Demarque, P., Kim, Y.-C., Lee, Y.-W., Ree, C. H., Lejeune, T., & Barnes, S. 2001, ApJS, 136, 417