MCR-TRGB: A Multiwavelength-covariant, Robust Tip of the Red Giant Branch Measurement Method

M. J. Durbin1, R. L. Beaton2, J. J. Dalcanton1, B. F. Williams1, and M. L. Boyer3
1 Department of Astronomy, University of Washington, Box 351580, U.W., Seattle, WA 98195-1380, USA; mjdurbin@uw.edu
2 Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA
3 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
Received 2020 January 31; revised 2020 June 12; accepted 2020 June 12; published 2020 July 23

Abstract
We present a new method to measure colors and magnitudes of the tip of the red giant branch (TGRB) in multiple bandpasses simultaneously by fitting an n-dimensional Gaussian to photometry of candidate tip stars. We demonstrate that this method has several advantages over traditional edge detection, particularly in regimes where the TRGB magnitude is strongly color dependent, as is the case in the near-infrared (NIR). We apply this method to a re-reduction of a set of optical and NIR Hubble Space Telescope data originally presented in Dalcanton et al. (D12). The re-reduction takes advantage of the increased depth and accuracy in the NIR photometry enabled by simultaneous reduction with higher-resolution optical data in crowded fields. We compare three possible absolute calibrations of the resulting apparent TRGB measurements, one adopting the same distance moduli as in D12 and two based on predicted TRGB absolute magnitudes from two widely used, modern sets of model isochrones. We find systematic offsets among the model absolute calibrations at the ~0.15 mag level, in line with previous investigations. The models also have difficulty reproducing the optical–NIR color–magnitude behavior of our measurements, making these observations a useful benchmark for future improvements.

Unified Astronomy Thesaurus concepts: Distance indicators (394); Red giant tip (1371); Near infrared astronomy (1093)

Supporting material: figure sets

1. Introduction
The tip of the red giant branch (TRGB) is defined as the truncation of the RGB phase of stellar evolution. The TRGB is reached when the helium flash ignites, terminating the expansion and cooling of the outer layers (Salaris & Cassisi 2006). Helium ignition occurs at a more or less fixed temperature, and thus the maximum bolometric luminosity ($L_{\text{bol}}$) produced by the core is well constrained (see, e.g., Sweigart & Gross 1978; VandenBerg et al. 2000; Salaris & Cassisi 2006; Serenelli et al. 2017). However, both the bolometric luminosity and the observed luminosity in a given bandpass will vary from star to star depending on the effective temperature, on the atmospheric chemistry, and on which elements and molecules selectively absorb and emit flux. While the TRGB can be used as a “standardizable candle,” care must be taken to understand the wavelength dependence of the observed TRGB luminosity (see Serenelli et al. 2017 for a discussion of additional physical details).

Baade (1944), when first resolving M31 into stars, noticed a field of red stars of roughly equal brightness, which we now associate with the TRGB of “old” stellar populations. However, the optical TRGB (OPT-TRGB) was not used as a distance indicator until Lee et al. (1993), who leveraged precise color–magnitude diagrams (CMDs) of globular clusters from Da Costa & Armandroff (1990) to demonstrate an effective technique to “detect” the truncation of the RGB sequence observationally and thereby determine a distance to the host system.

The Lee et al. (1993) methods are conceptually simple; to detect the truncation of the RGB sequence, one identifies the magnitude at which there is a sharp jump in star counts, as expected for the edge of the RGB sequence. Lee et al. (1993) applied an edge detection algorithm that approximates the first derivative of a discrete function (a Sobel filter, Sobel & Feldman 1968) to measure the point of greatest change in the RGB luminosity function (LF), which they identified as the apparent magnitude of the TRGB. Since Lee et al. (1993), algorithms to detect the TRGB and calibrations of the absolute TRGB have evolved (a review and comparison are given in Beaton et al. 2018), but the core of the technique has stayed the same. In general, the OPT-TRGB employed in the $I$ filter is thought to have a near-constant magnitude $M_I \sim -4$ mag for most old ($t > 5$ Gyr) and metal-poor ([Fe/H] $<-0.5$ dex) stellar populations—populations that are nearly ubiquitous in galaxies of all Hubble types and luminosity classes (Kunder et al. 2018). These properties have made the detection of the OPT-TRGB an effective distance determination method out to $\sim 31$ Mpc (Jang & Lee 2017b).

While the OPT-TRGB is a powerful tool with several key science drivers (e.g., Tully et al. 2013; Anand et al. 2019b; Freedman et al. 2019; Trujillo et al. 2019, among others), extending this method to the near-infrared (IR-TRGB hereafter) has several advantages: (i) the stars themselves are $\sim 1–1.5$ mag brighter and comparable in luminosity to $P \sim 10$ day Cepheids (see Figure 30 in Beaton et al. 2018); (ii) the impact of extinction is reduced by up to a factor of 6 (Indebetouw et al. 2005; Casagrande & VandenBerg 2014), permitting exploration of galaxies behind high extinction (see, e.g., Anand et al. 2019a) and reducing any dust-based systematics; and (iii) the next generation of astronomical facilities, whether 30 m class...
telescopes on the ground, wide-field telescopes in space, or large-aperture telescopes in space, are likely to realize their highest efficiency in the near- to mid-infrared. Thus, there is enormous potential for the IR-TRGB, although there remain challenges to its implementation at high precision.

The first detailed characterization of the IR-TRGB in the Hubble Space Telescope (HST) WFC3/IR filters was presented in Dalcanton et al. (2012a, D12 hereafter), in which 23 galaxies with optical imaging from the Advanced Camera for Surveys (ACS) Nearby Galaxy Survey Treasury (Dalcanton et al. 2009, hereafter D09) were supplemented with WFC3/IR imaging in the F110W and F160W filters (Dalcanton 2009, GO-11719). D12 detected the IR-TRGB applying a Sobel filter to F110W–F160W CMDs and then converted the dust-corrected apparent magnitudes to an absolute scale via distances derived in D09, using the OPT-TRGB calibrated to models described in Girardi et al. (2008). D12 found a strong correlation between the absolute F160W magnitude of the IR-TRGB and the F110W–F160W color, such that redder TRGB stars had a brighter absolute magnitude. The correlation was expected owing to metallicity variations among the sample, such that more metal-rich stars had redder colors, pushing a larger fraction of their bolometric flux into the near-IR. However, the D12 IR-TRGB was brighter than contemporaneous theoretical models by 0.05–0.10 mag and, generally, was notably different from globular cluster observations that had been converted from the Two Micron All Sky Survey (2MASS) into the WFC3/IR system. The general conclusion from this paper was that while the IR-TRGB was promising, there were significant unresolved issues. A subsequent and similar analysis by Wu et al. (2014), however, essentially found the same underlying magnitude–color relationship for the IR-TRGB, although these authors argued for a break in the slope at F110W–F160W = 0.95 mag.

More recent, ground-based work in the 2MASS filter system by Hoyt et al. (2018), Madore et al. (2018), and Görski et al. (2018) produced empirical color–magnitude relations for the IR-TRGB. These, however, are significantly different from those determined for WFC3/IR on HST. In their review, Beaton et al. (2018) compared the WFC3/IR and 2MASS IR-TRGB slopes to demonstrate that these independent WFC3/IR and 2MASS calibrations largely agree when considered within a given filter system and that the apparent differences are more likely due to inherent differences between the filter systems. As a result, calibrations from the ground-based 2MASS systems are likely inapplicable to the space-based WFC3/IR system.

In addition to advancing empirical measurements of the IR-TRGB, recent papers have also updated theoretical relationships derived from stellar models. A key work is that of Serenelli et al. (2017), which directly compared the theoretical IR-TRGB for a range of metallicities and ages in the BaSTI (Pietrinferni et al. 2013) model suite. Serenelli et al. (2017) report both physical and color–magnitude relationships for the IR-TRGB but note that uncertainties in the bolometric corrections and stellar $T_{\text{eff}}$ scale make direct use of these relationships challenging (as discussed further in Beaton et al. 2018). McQuinn et al. (2019) studied the variation in the TRGB with age and metallicity from the optical to the mid-IR using simulated photometry based on the PARSEC (Bressan et al. 2012; Marigo et al. 2017) model suite and found that rectifying the photometry to a fiducial tip reduced the range of variations in the measured F160W TRGB to 0.04 mag. Thus, while the potential for the IR-TRGB is well recognized (see, e.g., Beaton et al. 2018, among others), the empirical evidence for its reliability remains unclear.

D12 presented a number of concerns regarding their analysis that ranged from the relatively new data processing and calibration of WFC3/IR data to crowding in the images (for which the higher-resolution optical images are clearly deeper and more complete). However, since D12, major large-scale projects like the Panchromatic Hubble Andromeda Treasury (Dalcanton et al. 2012a; Williams et al. 2014), the Cosmic Assembly Near Infrared Deep Extragalactic Legacy Survey (Grogin et al. 2011; Koekemoer et al. 2011), and the Ultra Deep Field (Koekemoer et al. 2013; Borlaff et al. 2019) have led to substantial improvement both in our technical knowledge of the WFC3/IR camera and in the development of multi-wavelength data-processing techniques that significantly improve the WFC3/IR photometric quality. Additionally, there have been multiple internal efforts to improve WFC3/IR calibration and data products (for a comprehensive overview see Mack 2018). It is the purpose of this work to apply these techniques to the D12 data set and revisit the discrepancies identified in D12 regarding the IR-TRGB (Durbin 2017). We also take advantage of and expand on recent works (e.g., Hoyt et al. 2018; Madore et al. 2018; Freedman et al. 2020) that have demonstrated the effectiveness of calibrating the TRGB in multiple bandpasses by selecting a set of fiducial “tip stars” and fitting their multiwavelength behavior; we present a generalized version of this method here.

The outline of the paper is as follows. Section 2 describes the observations, image processing, photometry, and artificial-star tests (ASTs). Section 3 presents techniques to isolate the RGB, identify candidate TRGB stars, and trace their multicolor behavior. Section 4 presents the final measured TRGB apparent magnitudes and colors and compares the absolute magnitudes and distance moduli obtained from previously published distances and then from calibration to two sets of theoretical isochrones. Section 5 presents a discussion of our results, attempts to resolve concerns from D12, and discusses lingering concerns regarding the full realization of the IR-TRGB. Section 6 presents a summary of our work and discusses directions of future research. Throughout the main text, we limit visualizations to a representative set of galaxies; figures for the full sample are given as figure sets.

2. Data

2.1. Observations

We re-reduced the optical and near-infrared HST imaging data described in D09 and D12. The D12 observations were a WFC3/IR imaging follow-up (SNAP-11719) to the optical ACS/WFC data presented in D09. The F110W–F160W observations cover 26 pointings in 22 Local Volume galaxies with a range of star formation histories. The majority of the galaxies are low-metallicity dwarfs, with the exception of M81. Table 1, reproduced from D12, presents summary information about the galaxies in our sample, including coordinates, angular diameter, apparent $B$ magnitude, foreground reddening, T-type, H1 line widths, and group membership. We note that not all of these galaxies have the purely old stellar populations that are considered optimal for measuring the TRGB.

We analyzed 24 of the 26 data sets that were included in D12. To maintain uniformity in the final data set and analyses, we excluded two targets (NGC 404 and NGC 2403-DEEP) because
their optical data were taken by WFPC2 rather than ACS. Additionally, we combined the two pointings of Holmberg II because their optical data were taken by WFPC2 rather than ACS. To address this issue, we used the ACS to improve the alignment of the images. This was achieved by aligning the ACS with the WFPC2 images using TweakReg, a tool designed for astrometry.

Table 2 describes the ACS/WFC and WFC3/IR observations used for this work, including references to the original proposals, total F814W exposure time, and offsets of the observation from the galaxy center. Table 3 presents the ACS/WFC and WFC3/IR observations used for this work, including references to the original proposals, total F814W exposure time, and offsets of the observation from the galaxy center.

2.2. Alignment and Photometry

We aligned all exposures using TweakReg and Drizzle-pac 2.0 (Hack et al. 2013; Avila et al. 2015). TweakReg aligns images by calculating an affine transform (shifts, rotation, and scale) that best describes the transformation between astrometric catalogs from two images, one of which is treated as the fiducial “reference” image. It then calculates an updated WCS solution for the nonreference image using the affine transform.

By default, TweakReg extracts astrometric source catalogs from input images with a point-source extraction routine based on DAOFAST (Stetson 1987), which is optimized for point-source detection. However, many of our exposures are too sparsely populated with bright stars to produce a reliable cross-filter alignment solution from point sources alone, requiring the addition of background galaxies to the astrometric source catalogs. We therefore followed the procedure described in Lucas (2015) to align images on Source Extractor (Bertin & Arnouts 1996) catalogs rather than TweakReg-produced catalogs. Source Extractor’s detection algorithm is largely morphology-agnostic, which enables the robust detection of both point and extended sources. We used SEP (Barbary 2016), a Python and C reimplementation of core Source Extractor algorithms, to derive all catalogs used in alignment.

We chose ACS/WFC F814W as our “reference” filter for all targets, as it is the only optical filter common to all targets, and in most cases it is the deepest and most likely to contain sources that are detected across multiple filters. We aligned all frames for each target with the following steps:

1. Extract initial source catalogs from all F814W exposures with SEP and align these with TweakReg.
2. Combine all aligned F814W exposures into a single distortion-corrected reference image with AstroDrizzle and extract a deep reference catalog from the drizzled image.
3. Realign all F814W exposures to the reference image using catalogs from the cosmic-ray-cleaned (*crcllean) images produced by AstroDrizzle.

Note. Reproduced from D12, with updates to $A_V$ from Schlafly & Finkbeiner (2011). Name, position, diameter, $B_V$, $W_{50}$, and $T$-type taken from Karachentsev et al. (2004). $m − M$ from D09 and Karachentsev et al. (2003) for NGC 7793; group membership from Karachentsev (2005) or Tully et al. (2006).
4. Align all other exposures to the reference image with TweakReg.

We did not attempt to derive an absolute astrometric solution for any of our targets, as the majority are severely limited by the $\sim 2\' \times 2\'$ WFC3/IR field of view and do not have enough bright sources to reliably match against external astrometric catalogs such as Gaia. For the purposes of this work, internally consistent alignment on a per-target basis is sufficient.

Figure 2 compares the rms scatter of the alignment residuals for the common filters of F814W, F110W, and F160W. The residuals for the two WFC3/IR filters are very similar, with a residual scatter of $\sim 0.025$ or 0.2 WFC3/IR pixels. The residuals for F814W are more scattered, with a peak at 0.01 (0.2 ACS/WFC pixels) and a long tail, likely due to differences in the underlying image data sets themselves (e.g., different exposure depths and source densities).

We carried out photometry on the aligned images with the pipeline described in Williams et al. (2014), which wraps the HST photometry package DOLPHOT (Dolphin 2000). Briefly, DOLPHOT uses a set of fiducial point-spread function models that are empirically scaled for each frame to account for frame-to-frame differences, such as those induced by “breathing” (Hasan & Bely 1994). The cross-camera wrapper utilizes a single underlying source list such that DOLPHOT can iteratively measure each individual source simultaneously across frames employing techniques optimized for crowded fields. As described in Williams et al. (2014), the output photometry for each source requires additional characterization to have realistic uncertainties incorporating all concerns; these are derived via ASTs that are described in the following subsection.

The key difference in the procedure adopted here compared to that of D12 is that we perform simultaneous cross-camera photometry rather than reducing the data sets independently and then matching cataloged sources. Due to the differences in the native angular resolution between ACS/WFC and WFC3/IR (0.05 pixel$^{-1}$ vs. 0.13 pixel$^{-1}$ respectively), the simultaneous procedure should produce a more complete and robust WFC3/IR data set owing to improved deblending and more complete source lists.

We rejected large contaminating sources, such as bright foreground stars and background galaxies, by convolving the images with a 2D Gaussian kernel with width 0.75 (15 WFC3/IR pixels) and extracting sources from the convolved images with SEP. We used the ellipse parameters $a$, $b$, and $\theta$ of the sources to mask potentially contaminated pixels, with $a$ and $b$ multiplied by 5 to ensure that a sufficient fraction of the contaminating flux was masked.

2.3. ASTs

The primary sources of photometric uncertainty in these data are total exposure depth, which determines the Poisson noise of photon counts, and stellar surface density, which affects the likelihood of a star being blended with surrounding sources. The former are well captured by DOLPHOT’s accounting of photon-counting uncertainties. The latter, however, require
| Galaxy   | Target Name | Date Obs.       | Offset (arcmin) | Exptime (s) | $\sum_{\text{min}}$ | $\sum_{\text{max}}$ | $N_{\star}$ | Opt. Propid | Opt. Filters |
|----------|-------------|----------------|----------------|-------------|-----------------|----------------|-----------|-------------|--------------|
| KDG 63   | DDO 71      | 2010 Apr 21 16:33:04 | 0.97           | 9000        | 0.00            | 0.86            | 68477     | GO-9884     | F606W, F814W |
| DDO 78   | DDO 78      | 2010 Apr 20 15:13:25 | 0.34           | 2292        | 0.02            | 0.54            | 56458     | GO-10915    | F475W, F814W |
| DDO 82   | DDO 82      | 2010 May 7 07:27:41 | 0.38           | 2442        | 0.01            | 2.99            | 187699    | GO-10915    | F475W, F606W, F814W |
| KDG 2    | ESO 540-030 | 2009 Dec 17 12:32:10 | 0.15           | 7840        | 0.00            | 0.60            | 28087     | GO-10503    | F606W, F814W |
| HS 117   | HS 117      | 2010 Feb 24 02:35:38 | 0.13           | 900         | 0.00            | 0.46            | 13011     | GO-9771     | F606W, F814W |
| I2574    | IC 2574-SGS | 2010 Feb 25 03:34:37 | 3.28           | 6400        | 0.10            | 1.40            | 286852    | GO-9755     | F555W, F814W |
| KDG 73   | KDG 73      | 2010 Jun 9 18:17:42 | 0.43           | 2274        | 0.00            | 0.22            | 12721     | GO-10915    | F475W, F814W |
| KKH 37   | KKH 37      | 2009 Sep 29 11:12:38 | 0.09           | 3441        | 0.00            | 1.36            | 30966     | GO-10915, GO-9771 | F475W, F606W, F814W |
| M81      | M81-DEEP    | 2010 Jun 13 01:26:19 | 13.88          | 2953        | 0.02            | 0.23            | 63093     | GO-10915    | F475W, F606W, F814W |
| N300     | NGC 0300    | 2010 Apr 19 18:17:32 | 6.26           | 2982        | 0.05            | 0.59            | 197750    | GO-10915, GO-9492 | F475W, F555W, F606W, F814W |
| N2403    | NGC 2403-HALO-6 | 2010 Apr 25 04:57:54 | 5.58           | 720         | 0.01            | 0.46            | 20441     | GO-10523    | F606W, F814W |
| N2976    | NGC 2976-DEEP | 2010 Feb 25 02:34:59 | 3.03           | 27191       | 0.02            | 1.30            | 96662     | GO-10915    | F475W, F606W, F814W |
| N3077    | NGC 3077-PHOENIX | 2010 Feb 21 23:20:39 | 3.89           | 19200       | 0.02            | 0.38            | 70482     | GO-9381     | F555W, F814W |
| N3741    | NGC 3741    | 2009 Nov 7 02:03:02 | 0.51           | 2331        | 0.00            | 2.14            | 48369     | GO-10915    | F475W, F814W |
| N4163    | NGC 4163    | 2010 Mar 23 18:11:32 | 0.23           | 3150        | 0.01            | 3.89            | 153523    | GO-10915, GO-9771 | F475W, F606W, F814W |
| N7793    | NGC 7793-HALO-6 | 2010 Jun 14 19:43:15 | 6.02           | 740         | 0.01            | 0.45            | 20079     | GO-10523    | F606W, F814W |
| Sc22     | SCL-DE1     | 2009 Sep 08 01:16:49 | 0.18           | 17920       | 0.00            | 0.24            | 18967     | GO-10503    | F606W, F814W |
| N2403    | SN-NGC 2403-PR | 2010 Apr 22 08:27:47 | 0.90           | 1450        | 0.37            | 7.59            | 433196    | GO-10182, GO-10402 | F475W, F606W, F814W |
| HoII     | UGC 4305    | 2010 Feb 26 10:10:22 | 0.54           | 9920        | 0.02            | 1.65            | 327523    | GO-10605, GO-10522 | F555W, F814W |
| DDO 53   | UGC 4459    | 2010 Apr 23 11:46:34 | 0.25           | 4768        | 0.02            | 0.47            | 63451     | GO-10605    | F555W, F814W |
| Hol      | UGC 5139    | 2009 Aug 21 23:20:49 | 0.35           | 5936        | 0.04            | 0.52            | 105305    | GO-10605    | F555W, F814W |
| U8508    | UGC 8508    | 2009 Oct 14 20:11:32 | 0.18           | 2349        | 0.00            | 1.57            | 73755     | GO-10915    | F475W, F814W |
| UA 292   | UGCA 292    | 2010 May 18 13:08:17 | 0.35           | 2274        | 0.00            | 0.21            | 17668     | GO-10915, GO-10905 | F475W, F606W, F814W |

**Note.** Here the date observed is the date of the last IR exposure; the offset is the distance between the center of the IR footprint and the galaxy coordinates as given in Table 1; the exposure time is the total F814W exposure time (all IR observations have uniform exposure times of 600 s in F110W and 900 s in F160W); and $N_{\star}$ is the number of stars that were detected in all of F814W, F110W, and F160W.
additional tests to fully characterize, especially given that blending is typically the dominant source of bias and uncertainty in crowding-limited data.

We evaluated the photometric biases, scatter, and completeness of our data with a series of ASTs. We generated 20,000 artificial stars to be injected into the image stack for each target, for a total of 460,000 ASTs. We prioritized the near-IR RGB when selecting artificial-star magnitudes. Half were drawn directly from simulated absolute photometry generated with MATCH (Dolphin 2002) from PARSEC models. The other half were assigned random magnitudes within our F110W–F160W selection box, with optical magnitudes taken from simulated stars with comparable near-IR photometry. Figure 3 shows a CMD of the full set of input near-IR photometry. All absolute input magnitudes were then adjusted by the per-filter foreground reddening and D12 distance modulus for each target and assigned random pixel coordinates within the near-IR image footprints, excluding the locations of masked contaminating sources such as extended background galaxies and bright foreground stars. These input stars were then inserted into the image stack in batches of 1000 at a time and processed identically to the original photometry.

As the AST input locations were assigned at random, they do not necessarily reflect the true distributions of density and depth for any single target. We therefore resampled the full set of AST results to match the distribution of these quantities for each target as closely as possible, as follows.

We evaluated stellar surface densities using kernel density estimation (Rosenblatt 1956; Parzen 1962) with the Python package KDEpy (Odland 2018). We selected the photometry to be used for density estimation using the same near-IR selection box as in the ASTs, with the additional criteria of having a mean near-IR signal-to-noise ratio greater than 3. We then constructed stellar surface density maps by convolving source coordinates with a Gaussian kernel with a width of 5″ and tagged all photometry with their local densities. Density maps for three example targets are shown in Figure 4. In the analysis presented in Section 3 we used only photometry with local densities less than 1.5 stars arcsec$^{-2}$, except for the high-density target SN-NGC 2403-PR, where we used a maximum local density of 3 stars arcsec$^{-2}$.

While all near-IR exposures were taken with identical exposure times and are therefore of comparable depth, there is considerable variation in the optical exposure depths, which in turn may affect DOLPHOT’s source detection and subsequent deblending of near-IR sources. To characterize exposure depth, we use the weight maps generated by Astrodrizzle for the combined F814W reference images to assign fiducial total exposure times to the locations of each source.

For each target we separated the photometry into 10 bins according to density versus depth using K-means clustering (Arthur & Vassilvitskii 2007; Sculley 2010) and resampled the full set of ASTs to match the observed distributions of densities and depths.

We then use the resampled ASTs to assign fiducial photometric uncertainties, biases, and completenesses to all of our photometry. We define the photometric bias to be the median of the differences between observed and input AST magnitudes, the photometric error to be the interquartile range of the same, and the completeness to be the fraction of stars with non-null observed magnitudes. We calculate these quantities as a function of AST input magnitudes in each filter.

We subtract filter-appropriate foreground extinctions from all photometry, with values obtained from Schlafly & Finkbeiner (2011); the corresponding V-band extinctions are listed in Table 1. We assume negligible internal extinction for all targets, as the majority of our targets are either low-metallicity dwarfs or halos of spiral galaxies. Target SN-NGC 2403-PR is an exception, but in that case we find that the photometric uncertainties due to crowding are large enough that an attempt to analyze or correct for internal extinction would likely not be productive.

2.4. Comparison to D12 Photometry

Here we directly compare this generation of photometry to that of D12 by cross-matching individual stars. We first convert the IR pixel coordinates of the original photometry to the WCS defined by our realigned images. We select an initial sample of stars within 1 mag of the D12 TRGB values and maximum
filter old-to-new magnitude differences of 0.5 mag and match on R.A. and decl. using a kd-tree (Bentley 1975) with a maximum distance of 2″. We then find the robust coordinate transformation parameters between the new and old photometry using RANSAC regression (Fischler & Bolles 1987) on the matched initial sample with a maximum residual value of 0.1. We apply this transformation to the full set of old photometry coordinates and match the transformed coordinates again with a kd-tree, this time with a maximum distance of 0.1. Figure 5 shows the changes in magnitude as a function of the original D12 magnitudes in F160W, with the D12 TRGB ±0.1 mag highlighted. Interestingly, we find that near the tip the median magnitude difference is typically very small (on the order of 0.01 mag) but negative for uncrowded stars, indicating that this generation of photometry is slightly brighter than the previous one. However, even the sparsest fields show a population of high-crowding stars that are several tenths of a magnitude dimmer than their D12 counterparts.

Figure 4. Left: Near-IR Hess diagrams of three galaxies in our sample, with the selections of stars included in our surface density calculations highlighted. Right: corresponding stellar surface density maps for each target. All density maps are scaled to the same limits (0–1.5 RGB stars per square arcsecond) to highlight the range of stellar densities in our sample. Gaps in the density images show where contaminating sources such as foreground stars and background galaxies were rejected. The complete figure set (23 images) is available in the online journal.

Figure 5. Changes in photometry between D12 and this work for HS 117 (top), NGC 300 (middle), and NGC 4163 (bottom), with the D12 magnitude on the x-axis and Δm on the y-axis. The color-coding indicates the DOLPHOT crowding parameter of the new photometry, which is the number of magnitudes subtracted from the initial measurement due to neighboring sources. The rolling mean and median are shown by the solid and dashed lines, respectively. The complete figure set for both F110W and F160W (46 images) is available in the online journal.

The Astrophysical Journal, 898:57 (26pp), 2020 July 20 Durbin et al.

3. TRGB Measurement

In this section we describe the steps we use to measure the apparent magnitudes and colors of the IR-TRGB. We adopt a multiwavelength approach, which we call “MCR-TRGB,” that we summarize for the reader in advance of detailed descriptions. First, we isolate the RGB sequence from the other stellar populations. From the RGB sample, we do a tip detection to select stars in the vicinity of the TRGB. This initial sample is
then separated into potential subpopulations to isolate those that have colors and magnitudes consistent with being TRGB stars. The color and magnitude distributions of the candidate tip stars are then fitted jointly for all applicable color–magnitude spaces to build the final color–magnitude calibrations. This approach has several advantages over traditional Sobel edge detection for the purpose of this work, which we discuss in detail in Section 5.

Throughout this section, the methods are demonstrated using galaxies that span a range in metallicity and RGB shape. Identical figures for each of the 23 galaxies in the sample are provided as figure sets.

### 3.1. Initial RGB Star Selection

A maximally complete and minimally contaminated sample of RGB stars is essential for characterizing the TRGB and the RGB LF near the tip. Unfortunately, there are many stars that have colors and magnitudes similar to RGB stars, such as red helium-burning and asymptotic giant branch (AGB) stars. These “contaminant” populations can blur the TRGB edge or distort its measured magnitude (see discussion in D12).

Typically, RGB stars are selected using strict binary color–magnitude cuts; we describe two particular examples. D12 initially select stars with colors in the range $0.6 \text{ mag} < F110W - F160W < 1.1 \text{ mag}$ and magnitudes brighter than 1 mag below their initial TRGB estimate and then make further rejections based on the standard deviations of a linear fit to the remaining stars in color–magnitude space. The Carnegie-Chicago Hubble Program (Hatt et al. 2017, 2018a, 2018b; Jang et al. 2018; Freedman et al. 2019) makes color cuts with a fiducial RGB slope and a color width chosen visually to encompass the edges of the RGB near the tip.

Here, we leverage the multiwavelength information available for our targets to probabilistically identify stars that fall along characteristic RGB color–magnitude sequences. For each target we construct a set of red versus blue–red CMDs using F814W, F110W, and F160W as the red filters, and using all available optical filters other than F814W as the blue. We also construct CMDs in F814W versus F814W–F160W, F110W versus F814W–F110W, and F110W versus F110W–F160W for all targets. The number of unique color–magnitude combinations varies from 6 to 12 depending on the number of available optical filters for each target. We apply broad initial color and magnitude cuts based on the D12 TRGB measurements. Figure 6 provides example CMDs after cuts for NGC 4163 in the color–magnitude combinations used for this analysis.

Next, we define an RGB locus in each filter combination by fitting a predicted RGB color–magnitude sequence to the photometry in each color–magnitude combination independently. We minimize the median distance between the observed photometry and a grid of synthetic photometry derived from PARSEC (Marigo et al. 2017) isochrones of ages 4–14 Gyr and [Fe/H] $-3.0$ to $-0.2$ dex, which have been limited to RGB stars brighter than $M_{F110W} = -2$ mag and converted to apparent magnitudes using the distance moduli from D12. The panels of Figure 6 have their best-fit isochrone-predicted RGB sequences overlaid in blue. As our goal is to trace the RGB color–magnitude locus across all available bandpasses rather than to measure any underlying properties of the stellar populations, we do not force a single age and metallicity combination to fit all color–magnitude combinations. (We note that the metallicities of the “best-fit” isochrones for a single target can vary from filter to filter by up to nearly a full dex, especially in the case of low-metallicity targets, where the upper RGB color only weakly depends on metallicity; see Section 5.3 for further discussion of filter-to-filter differences between observed and predicted photometry at the TRGB.)

An initial “RGB sequence probability” is then assigned to each star based on the distance between its observed position in color–magnitude space and the nearest point on the predicted RGB for each color–magnitude combination. The circles in the panels of Figure 6 are color-coded by these probabilities.

The purpose of this process is to construct an LF with which to make an initial TRGB estimate, as described in Section 3.2. We therefore extrapolate the fitted RGB sequences out to at least 1.5 mag brighter than the measured D12 TRGB apparent magnitudes in all filters. As a result, stars brighter than the TRGB that fall along the predicted color–magnitude loci will be assigned high RGB sequence probabilities. These probabilities should be understood as estimates of a star’s proximity to the color–magnitude relations characteristic of each target’s RGB sequence, rather than as identifications of only the stars that are truly on the RGB.

The individual RGB sequence probabilities are then averaged across all color–magnitude combinations to produce global RGB sequence probabilities, which we call $P(\text{RGB})$.

### 3.2. Edge Detection

We make an initial selection of candidate tip stars by applying a Sobel edge detection to the RGB-weighted LF. For each target we choose the filter with the sharpest LF, that is, the filter in which the tip magnitude is least dependent on color. This is F814W for most targets and F110W for targets with F110W – F160W $> 0.95$ mag, as measured in D12. We first construct an LF (shown in the middle column of Figure 7) by marginalizing $P(\text{RGB})$ over color as a function of magnitude. We use a bin size of 0.01 mag, which is a factor of $\sim 5$ smaller than the typical magnitude uncertainty.

For each galaxy, the middle panels of Figure 7 show the raw LF (blue), where the noise is consistent with Poisson fluctuations. We first smooth the LF with a Savitzky–Golay filter (Savitzky & Golay 1964), a low-pass filter originally developed to suppress noise in spectroscopic data by fitting a polynomial within a rolling window. This technique effectively removes Poisson noise spikes while preserving sharp features such as the TRGB edge. However, there may be remaining spurious edges from photometric variance or stochastic sampling of the LF, particularly in sparse data. To reduce the impact of these false edges, we smooth the LF once more using Gaussian-windowed, Locally Weighted Scatterplot Smoothing (GLOESS); for an in-depth description see Hatt et al. (2017) and references therein. Briefly, GLOESS is an implementation of one-dimensional Gaussian kernel density estimation, which we have modified to accept a variable kernel width. We select a fiducial kernel width using the KDEpy implementation of the Improved Sheather–Jones algorithm (Botev et al. 2010), which chooses an optimal kernel width based on the overall density of the data. We then multiply this fiducial width by the square of the photometric uncertainties scaled by their median value as a function of magnitude, which de-emphasizes LF variation fainter than the TRGB, where photometric uncertainties are higher. The final smoothed LF, shown overlaid in black on the raw LFs in Figure 7, is then used for the initial TRGB detection.
To detect the TRGB, we begin by applying a Sobel filter, which is one of the most widely used means of finding the tip (see summary and comparisons in Beaton et al. 2018). The Sobel filter approximates the first derivative of a discrete data set via convolution with a kernel. In its simplest form, this kernel is $[-1, 0, 1]$, which effectively subtracts counts in the $i-1$ bin from the $i+1$ bin to determine the edge response, $\eta$, for bin $i$. This kernel is applied to the smoothed LF, and the response is shown for each galaxy in the right panels of Figure 7. In Figure 7, the magnitude of maximum Sobel response, $m(\eta_{\text{max}})$, is indicated by the dashed line across all panels.

We then select candidate tip stars near $m(\eta_{\text{max}})$ within a range we call $\Delta \eta$. The value of $\Delta \eta$ is determined using two quantities: (i) the median photometric error within $\pm 0.1$ mag of $m(\eta_{\text{max}})$, $\sigma_{\text{phot}}^{\eta_{\text{max}}}$, and (ii) a minimum number of tip candidate stars $N_s^{\text{min}}$. We define $N_s^{\text{min}}$ as the square root of the number of stars 1 mag below $m(\eta_{\text{max}})$, with a hard minimum of 30 stars. For each target, we make an initial selection of stars within $\pm 1\sigma_{\text{phot}}^{\text{TRGB}}$ and then iteratively expand the selection range by 0.5$\sigma_{\text{phot}}^{\text{TRGB}}$ on each side until either $N_s^{\text{min}}$ is reached or $\Delta \eta$ is over 0.2 mag. For the majority of our targets, the initial selection window of $\pm 1\sigma_{\text{phot}}^{\text{TRGB}}$ is enough to meet $N_s^{\text{min}}$. Our final $\Delta \eta$ is shown by the blue band in the panels of Figure 7 for our example pointings.

Out of the stars that fall within the fiducial tip magnitude range, we first select likely RGB stars as those with $P(\text{RGB}) > 0.6$, which roughly corresponds to stars that were identified as RGB+ sequence candidates with over 90% probability in at least two-thirds of the filter combinations we used to assign RGB probabilities. We then reject stars with anomalous magnitudes in at least one filter with Local Outlier Factor outlier detection (Breunig et al. 2000), which evaluates the relative isolation of points using $k$-nearest neighbors. We
take this trimmed sample of stars to be our final set of tip star candidates, which we then use to measure tip magnitudes and colors as described in the following section.

We note that this selection of likely RGB tip stars is performed based on the results of applying the Sobel filter to the filter where the tip is “flat” with color. The Sobel filter, by design, looks for a sharp edge in a one-dimensional distribution. Two-dimensional implementations of the Sobel filter exist but still require conversion of our CMDs into a binned form. Thus, application of the one-dimensional Sobel filter to a distribution that has magnitude—color behavior may not fully detect the true edge in the distribution. Lastly, where there is strong magnitude—color trend, because our colors are more imprecise than our magnitudes, the intrinsic slope can be distorted by the color spread in our data. Thus, in the next section we develop a method to utilize the tip stars we have just identified to trace the intrinsic TRGB slope across our set of color–magnitude combinations.

### 3.3. Multiwavelength Tip Fitting

We characterize the color and magnitude distributions of our candidate tip stars using Extreme Deconvolution (XDGM; Bovy et al. 2011), a modification of Gaussian mixture modeling that accounts for uncertainties in the input data. Specifically, we use XDGMM to fit a single six-dimensional Gaussian to the F814W, F110W, and F160W magnitudes and the F814W–F160W, F814W–F110W, and F110W–F160W colors of the tip star candidates. Although the underlying distribution of tip stars in this parameter space is not intrinsically Gaussian, we find that a single Gaussian is a reasonable approximation for the majority of our tip star samples. Additionally, for the faintest and sparsest of our targets, low star counts and photometric uncertainties on the same order as the width of the tip star selection windows do not allow us to place reasonable constraints on more complex models, such as multicomponent Gaussian mixtures. We discuss potential alternative modeling approaches in Section 5.2.

For the uncertainties we use as inputs to XDGMM, we divide each star’s individual photometric uncertainties by $P$ (RGB+), effectively weighting the input points by $P$(RGB+). We emphasize that XDGMM, as a tool, allows us to take into account these uncertainties and weights on the RGB+ likelihood to trace the tip in filters where the Sobel edge is less effective owing to color–magnitude slopes.

We take the means of the fitted distributions to be our final apparent tip magnitudes and colors. Results of these fits are shown for our sample galaxies in Figure 8, where we plot ellipses showing the 95% confidence regions of the XDGMM fits in three color–magnitude combinations. The width, height, and position angle of each ellipse are derived from two-dimensional slices of the full six-dimensional covariance matrix.

Potential systematic and statistical biases of this method are discussed in the Appendix; overall, we find that the results are comparable to those of edge detection in most cases.

### 4. Results

#### 4.1. Apparent TRGB Magnitudes and Colors

In this section we compare the TRGB apparent magnitudes and colors we have measured using the techniques developed in this paper to those used in D12. All revised apparent magnitudes and errors are reported in Table 3.

First, Figure 9 compares the change in apparent F160W magnitude and F110W–F160W color between this work (blue circles) and D12 (orange circles) for each target in our sample. The 68% confidence intervals are shown for our measurements and demonstrate that the difference between this work and D12 is almost always larger than our measurement uncertainties, although, as shown in the lower right, most are within the color–magnitude photometric error circle for an individual source at the tip.

The origin of these offsets can be determined by comparing the individual differences between the photometry. Figure 10 compares the relative change between the measurements of
this work and those of D12 for the F110W (x-axis) and F160W (y-axis). For both $\Delta m_{F160W}$ and $\Delta m_{F110W}$ (defined as this work minus D12), the median difference is approximately +0.05 mag; histograms are shown in Figure 10 on each axis. Interestingly, the offsets are highly correlated; in the top panel, a one-to-one line is shown as the black dashed line, with a fit to the results given as a blue solid line, and the 95% confidence interval (shown in the shaded region) encompasses the one-to-one line.

Figure 11 displays the F814W–F110W (left) and F814W–F160W (right) to F110W–F160W color–color diagrams for the results of this work (blue) compared to that of D12 (orange). Relative to D12, the measurements from this work move the color–color relations to the left in this diagram—bluer in F814W–F110W and F814W–F160W and slightly redder in F110W–F160W.

In Figure 11 we provide reference lines to highlight the color behavior, using a linear function for the D12 photometry and a logistic function for our new photometry. (We caution that these fitting relations should not be taken as physically meaningful.)
4.2. The TRGB Color–Absolute Magnitude Relation

To derive the color dependence of the near-IR TRGB absolute magnitude, we must adjust the apparent magnitudes in Figure 9 by the appropriate distance modulus for each galaxy.

We first present a revised near-IR color–absolute magnitude relation adopting the same distances as in D12 and then explore the use of the most up-to-date stellar models to derive revised distance moduli and absolute magnitudes.

4.2.1. Adopting D12 Distances

The distance moduli used in D12 were determined using the F814W TRGB, which enables a fully self-consistent study of the TRGB across bandpasses. With the exception of NGC 7793, these distances were originally published in Dalcanton et al. (2009), whereas the distance for NGC 7793 is from Karachentsev et al. (2003), which also uses the F814W TRGB. The absolute calibration of the F814W (ground-based I-band) TRGB at <5% precision is unclear (Jang & Lee 2017a; Beaton et al. 2018; Freedman et al. 2019, 2020; Reid et al. 2019; Yuan et al. 2019). Historically, it has been assumed to be a constant value of approximately $M_{\text{TRGB}}^{F814W} \sim -4.05$ mag (Lee et al. 1993; Salaris & Cassisi 1997). However, this magnitude is only anticipated to be robustly constant for uniformly old and metal-poor populations (Salaris & Cassisi 2006; Serenelli et al. 2017; Beaton et al. 2018); more specifically, [M/H] > −0.5 dex and >4 Gyr.

The stellar populations in the D12 sample, however, span a wide range of ages and metallicities that preclude the assumption of a single value for the TRGB F814W luminosity. Rather than adopting a single value for $M_{\text{TRGB}}^{F814W}$, Dalcanton et al. (2009) used the mean optical colors of stars within 0.2 mag of the apparent F814W TRGB to choose fiducial Girardi et al. (2008) isochrones with corresponding colors. Dalcanton et al. (2009) then determined the predicted F814W TRGB absolute magnitude for each galaxy from the isochrone fitting, and the lower right error bars indicate the median photometric uncertainties in color and magnitude for an individual star. For NGC 300 (open circles) we plot $m_{\text{TRGB}}^{F160W}$ rather than $M_{\text{TRGB}}^{F160W}$, as it is ~1 mag brighter than the remainder of the sample. On average, our mean color–magnitude tip results are redder and slightly fainter than D12.

TRGB absolute magnitude for each galaxy from the isochrone sets, subtracted that from their measured F814W TRGB apparent magnitudes, and corrected for foreground extinction to obtain their distance moduli.

![Figure 9](image-url)
We calculate near-IR TRGB absolute magnitudes by subtracting the D12 distance moduli from our apparent tip magnitudes, as reported in Table 4. The resulting near-IR absolute magnitudes and color–magnitude relation are compared to those of D12 in Figure 12.

We fit a linear relation to the absolute F160W magnitudes and F110W–F160W colors determined in this work using orthogonal distance regression (Boggs et al. 1987) and find

$$M_{TRGB}^{F160W} = -2.541(F110W - F160W) - 3.475. \quad (1)$$

The uncertainties on our slope and zero-point are 0.057 mag color$^{-1}$ and 0.050 mag, respectively. Compared to the equivalent fit from D12 (their Equation (1)),

$$M_{TRGB}^{F160W} = -2.576(F110W - F160W) - 3.496, \quad (2)$$

we find an 0.02 mag fainter zero-point (<1% in distance) and a change in the slope of less than 0.04 mag color$^{-1}$, both of which are well within our uncertainties. As expected, the difference in the zero-point is roughly equivalent to the differences in measured TRGB photometry observed in Figure 9.

4.2.2. Recalibrating Distances to Recent Models

Both the physical isochrones and the filter transformations described in Girardi et al. (2008) have undergone many revisions in the intervening years (Bressan et al. 2012; Marigo et al. 2017), and thus the F814W TRGB zero-points adopted in D09 and D12 may no longer be appropriate. Here we apply a similar distance estimation method to that in D09 to our revised measurements, using synthetic photometry from the model suites PARSEC v. 1.2S5 (Bressan et al. 2012; Marigo et al. 2017) and MIST v. 1.26 (Choi et al. 2016), both of which are used routinely for stellar population work. We retrieved the synthetic photometry directly from the cited web services. For both sets, we use isochrones with ages spanning 8–14 Gyr with log(age) spacing of 0.05 dex. The PARSEC metallicities span $-2.2$ dex $\leq$ [Fe/H] $\leq$ 0 dex with a spacing of 0.1 dex, whereas the MIST metallicities span $-2.0$ dex $\leq$ [Fe/H] $\leq$ 0 dex with a spacing of 0.25 dex. Both model suites use scaled-solar abundances, albeit with slightly different calibrations ($Z_0 = 0.0152$ and $Y_0 = 0.275556$ for PARSEC, and $Z_0 = 0.0142$ and $Y_0 = 0.2703$ for MIST). We use the evolutionary phase tags in each model set to select the predicted TRGB at each age/metallicity combination.

For each of these model sets, we estimate new sets of distance moduli using two color–magnitude combinations: (i) F814W versus F814W–F160W and (ii) F160W versus F110W–F160W. For most of our color measurements, there are multiple isochrones with TRGB colors that fall within the measurement uncertainties, each with slightly different absolute TRGB magnitudes. We calculate a fiducial tip absolute magnitude for each galaxy by taking the weighted mean of the isochrone absolute magnitudes, where the weights are defined by a Gaussian with a center at the measured tip color and a width from the color uncertainty. Derived distance moduli for all color–magnitude combinations and model sets are reported in Table 5.

3 http://stev.oapd.imaf.fr/cgi-bin/cmd_3.3
6 http://waps.cfa.harvard.edu/MIST/index.html

The results of this procedure are shown in Figure 13, where the left panel shows the adopted values of $M_{TRGB}(F814W)$–(F814W–F160W) and the right panel shows the inferred values of $M_{TRGB}(F160W)$–(F110W–F160W). In each panel, sets of synthetic photometry at single ages (8–13 Gyr), with metallicities spanning $-2.0$ dex $\leq [Fe/H] < -0.25$ dex, are plotted as transparent lines, with PARSEC models plotted in green and MIST models in orange. Overall, the magnitude–color behavior of the two sets is qualitatively similar, but the absolute magnitudes differ by $\Delta M_{F814W} \sim 0.15$ mag, with
PARSEC being brighter than MIST for the same color (≈8% in distance).

The fits of our data to the predicted \( M_{\text{TRGB}}(F814W) \)-\( (F814W-F160W) \) distributions are shown as the circles in the panels of Figure 13. From this, we determine a distance modulus to each galaxy, which we denote as \( \mu \). In the right panel, we use \( \mu_{F814W} \) to translate \( m_{\text{TRGB}}(F160W) \) to \( M_{\text{TRGB}}(F160W) \) and compare these values to the same isochrone sets used to derive \( \mu_{F814W} \). The PARSEC-based distances place the observed near-IR TRGB ∼0.05 mag fainter than predicted, but they do trace the same underlying variations with color. In contrast, the MIST-derived values are less offset overall in magnitude but show shape deviations that become particularly pronounced at red colors.

Figure 14 repeats this process in reverse by determining a distance modulus, \( \mu_{F160W} \), based off of the \( M_{\text{TRGB}}(F160W) \)-\( (F110W-F160W) \) model predictions (left panel) and then comparing the absolute \( M_{\text{TRGB}}(F814W) \)-(\( F814W-F160W \)) empirical relationship to the models in the right panel. In this case, different behavior is observed: for both model sets, \( M_{\text{TRGB}}(F814W) \) from our data are too bright by ∼0.1 mag at the blue (low-metallicity) end, and there is a slight color offset of ∼0.05 mag.

Overall Figures 13 and 14 suggest that the isochrone predictions are inconsistent both with each other and, when used internally to predict multiband magnitudes, with our measurements. We will further discuss potential reasons for these apparent inconsistencies in Section 5.3.

### 4.2.3. New Distance Moduli Compared to D12

Figure 15 compares the distance moduli determined in the previous subsection to those from D12. The top panel compares the distances calibrated to F814W (the same filter in either case), and the bottom panel compares the distances from F160W. No difference (\( \Delta \mu = 0 \) mag) is indicated by the vertical dashed line. In both cases, the PARSEC-based calibration is systematically larger than in D12 by median values of 0.043(±0.069) mag in F814W and 0.126(±0.071) mag in F160W, corresponding to 2(±3)% and 6(±3)% greater distances, respectively. The MIST calibrations, on the other hand, are systematically smaller, with median differences of −0.096(±0.084) mag in F814W and −0.045(±0.063) mag in F160W (4(±4)% and 2(±3)% closer, respectively).

### 5. Discussion

In this section, we discuss the advantages and limitations of our adopted methods to trace the TRGB across multiple wavelengths. Once established, we then discuss more fundamental limitations to our investigation, which include knowledge of the absolute magnitude of the TRGB, details of the physical models underlying the isochrone suites, and possible systematics that are difficult to disentangle with the data at hand.

#### 5.1. Advantages of the MCR-TRGB Method

MCR-TRGB simultaneously measures the distributions of a pre-selected group of stars across an arbitrary number of color–magnitude combinations. As a result, the method ensures self-consistency in the measured color–magnitude behavior, as it is determined from the same underlying set of stars. In contrast, traditional edge detection is done on a per-filter basis, and it is not guaranteed to detect the tip using precisely the same stars across color–magnitude combinations. This limitation is particularly important in the cases of steeply sloped tips where the corresponding color baseline changes significantly relative to the color uncertainty.

Fitting the full color–magnitude covariance has the further benefit of characterizing spread of colors and magnitudes both...
Table 4
Absolute TRGB Magnitudes from D12 Distances

| Target             | $\mu$ (D12) | $M_{F814W}$ | $\sigma_{F814W}$ | $M_{F110W}$ | $\sigma_{F110W}$ | $M_{F160W}$ | $\sigma_{F160W}$ |
|--------------------|-------------|-------------|------------------|-------------|------------------|-------------|------------------|
| DDO 71             | 27.740      | -3.998      | 0.016            | -4.750      | 0.052            | -5.606      | 0.062            |
| DDO 78             | 27.820      | -4.090      | 0.033            | -4.854      | 0.065            | -5.764      | 0.078            |
| DDO 82             | 27.900      | -4.036      | 0.044            | -4.860      | 0.066            | -5.777      | 0.079            |
| ESO 540-030        | 27.610      | -3.993      | 0.037            | -4.670      | 0.046            | -5.518      | 0.053            |
| HS 117             | 27.910      | -4.065      | 0.037            | -4.756      | 0.045            | -5.592      | 0.044            |
| IC 2574-DEEP       | 27.900      | -4.025      | 0.040            | -4.809      | 0.069            | -5.689      | 0.076            |
| KDG 73             | 26.030      | -4.143      | 0.037            | -4.772      | 0.051            | -5.547      | 0.057            |
| KKH 37             | 27.560      | -4.018      | 0.032            | -4.741      | 0.049            | -5.601      | 0.056            |
| M81-DEEP           | 27.770      | -3.696      | 0.104            | -4.878      | 0.045            | -5.878      | 0.043            |
| NGC 0300           | 26.500      | -4.007      | 0.046            | -4.935      | 0.024            | -5.898      | 0.031            |
| NGC 2403-HALO-6    | 27.500      | -4.160      | 0.034            | -5.003      | 0.045            | -5.907      | 0.046            |
| NGC 2976-DEEP      | 27.760      | -4.026      | 0.061            | -4.903      | 0.047            | -5.850      | 0.049            |
| NGC 3077-PHOENIX   | 27.920      | -3.948      | 0.082            | -4.930      | 0.047            | -5.910      | 0.053            |
| NGC 3741           | 27.550      | -4.062      | 0.035            | -4.755      | 0.043            | -5.569      | 0.044            |
| NGC 4163           | 27.290      | -4.049      | 0.031            | -4.782      | 0.055            | -5.667      | 0.070            |
| NGC 7793-HALO-6    | 27.900      | -4.092      | 0.052            | -4.931      | 0.056            | -5.859      | 0.057            |
| SCL-DEI            | 26.110      | -4.103      | 0.031            | -4.762      | 0.060            | -5.556      | 0.060            |
| SN-NGC 2403-PR     | 27.500      | -4.084      | 0.101            | -5.043      | 0.105            | -6.011      | 0.086            |
| UGC 4305           | 27.650      | -4.081      | 0.034            | -4.847      | 0.060            | -5.696      | 0.069            |
| UGC 4459           | 27.790      | -4.082      | 0.034            | -4.805      | 0.052            | -5.628      | 0.057            |
| UGC 5139           | 27.950      | -4.057      | 0.022            | -4.817      | 0.056            | -5.654      | 0.069            |
| UGC 8508           | 27.060      | -4.042      | 0.027            | -4.745      | 0.039            | -5.557      | 0.049            |
| UGCA 292           | 27.790      | -4.040      | 0.035            | -4.622      | 0.049            | -5.379      | 0.060            |

![Figure 12](https://example.com/figure12.png)

**Figure 12.** Comparison of revised near-IR color–absolute magnitude relations to D12 with revised absolute magnitudes derived using the same distances as in D12. Blue circles are values from the current work, and orange circles are from D12. The corresponding color-coded lines show linear fits to each data set, and the shaded regions show 95% confidence intervals. Again, we see that our results are slightly redder and slightly fainter than D12 using their distances. This color–color relation is distance independent.

Across a given TRGB, and as a result, the Sobel-edges are generalized to a mean color on a per-filter basis.

Unlike the $T$-magnitude system of Madore et al. (2009), which rectifies the photometry to an assumed TRGB slope, MCR-TRGB relies only on the assumption that there is one filter in which the tip magnitude has a weak enough color dependence to make an initial selection of tip star candidates. This permits us to use the well-established Sobel method to find a “flat” edge to define candidate tip stars and then utilize those stars in regimes where the tip is more difficult to detect using Sobel-based methods. This provides a fundamental advantage toward revealing the underlying intrinsic behavior of TRGB stars to construct self-consistent color–magnitude calibrations. We posit that MCR-TRGB is a more effective tool to define the underlying color–magnitude calibrations for local, well-studied, and well-observed galaxies than relying on techniques more suited for distant galaxies. Stated differently, if your goal is to provide the best characterization of the behavior of tip stars across multiple bands, MCR-TRGB will perform better than standard Sobel edge detection. It also provides a fully empirical basis to explore the ultimate precision of the TRGB as a distance-measurement tool, from which we can gain understanding of both systematic and statistical biases for more distance measurements where there is a lower ability to probe these terms with available data.

5.2. Limitations of MCR-TRGB

Our method requires multiwavelength data and relatively good data quality, which makes it less generally applicable to all distance-measurement applications. More specifically, MCR-TRGB requires fairly stringent initial rejection of potential contaminants, which may not be feasible for all data sets owing to photometric uncertainties or the complexity of the underlying stellar populations. Moreover, the multiwavelength tracing of individual TRGB stars may also be infeasible for many contexts where the acquisition of multiband imaging is
PARSEC and MIST is present for the absolute magnitudes concerns for the isochrone sets:

Note. Quoted errors are the quadrature sum of the photometric and fitting errors.

| Target              | PARSEC       | MIST         |
|---------------------|--------------|--------------|
|                     | F814W | F160W | F814W | F160W |
|                      | µ     | σ     | µ     | σ     | µ     | σ     |
| DDO 71              | 27.842| 0.065 | 27.899| 0.093 | 27.710| 0.065 | 27.747| 0.093 |
| DDO 78              | 27.822| 0.093 | 27.939| 0.100 | 27.700| 0.093 | 27.785| 0.100 |
| DDO 82              | 27.948| 0.101 | 28.026| 0.109 | 27.835| 0.101 | 27.864| 0.109 |
| ESO 540-030         | 27.719| 0.072 | 27.839| 0.080 | 27.566| 0.072 | 27.694| 0.080 |
| HS 117              | 27.948| 0.072 | 28.042| 0.075 | 27.796| 0.072 | 27.885| 0.075 |
| IC 2574-SGS         | 27.969| 0.094 | 28.027| 0.115 | 27.847| 0.094 | 27.866| 0.115 |
| KDG 73              | 27.978| 0.072 | 28.077| 0.090 | 27.761| 0.072 | 27.898| 0.090 |
| KKH 37              | 27.643| 0.072 | 27.733| 0.087 | 27.506| 0.072 | 27.585| 0.087 |
| M81-DEEP            | 28.043| 0.162 | 28.068| 0.070 | 28.104| 0.064 | 28.188| 0.094 |
| NGC 0300            | 26.551| 0.079 | 26.639| 0.047 | 26.441| 0.079 | 26.435| 0.047 |
| NGC 2403-HALO-6     | 27.423| 0.061 | 27.456| 0.060 | 27.311| 0.061 | 27.306| 0.060 |
| NGC 2976-DEEP       | 27.804| 0.098 | 27.893| 0.079 | 27.696| 0.098 | 27.699| 0.079 |
| NGC 3077-PHOENIX    | 28.015| 0.124 | 28.118| 0.088 | 27.909| 0.124 | 27.877| 0.083 |
| NGC 3741            | 27.590| 0.065 | 27.659| 0.075 | 27.429| 0.065 | 27.501| 0.075 |
| NGC 4163            | 27.340| 0.079 | 27.450| 0.091 | 27.209| 0.079 | 27.290| 0.091 |
| NGC 7793-HALO-6     | 27.948| 0.091 | 28.026| 0.077 | 27.838| 0.091 | 27.858| 0.077 |
| SCL-DEI             | 28.104| 0.064 | 28.188| 0.094 | 27.926| 0.064 | 28.020| 0.094 |
| SN-NGC 2403-PR      | 27.467| 0.158 | 27.546| 0.155 | 27.358| 0.158 | 27.334| 0.155 |
| UGC 4305            | 27.668| 0.084 | 27.705| 0.099 | 27.537| 0.084 | 27.557| 0.099 |
| UGC 4459            | 27.810| 0.074 | 27.857| 0.087 | 27.664| 0.074 | 27.695| 0.087 |
| UGC 5139            | 27.993| 0.075 | 28.022| 0.099 | 27.858| 0.075 | 27.864| 0.099 |
| UGC 8508            | 27.119| 0.064 | 27.178| 0.069 | 26.964| 0.064 | 27.015| 0.069 |
| UGCA 292            | 27.811| 0.071 | 27.961| 0.091 | 27.586| 0.071 | 27.787| 0.091 |

Note. Quoted errors are the quadrature sum of the photometric and fitting errors.

---

too expensive. Thus, as just discussed, we consider that MCR-TRGB’s most significant role is as a tool to define the underlying systematics affiliated with TRGB-based distances as the community explores different color–magnitude regimes. The XDGMM algorithm used in MCR-TRGB loses its ability to resolve the shape of an intrinsic distribution when the typical uncertainties on the input data points are comparable to the full range of the input data. This method loses its advantages for low signal-to-noise photometry, filter combinations with very short color baselines, and simple stellar populations with little color spread near the TRGB. (See the Appendix for further discussion.) Another limitation of XDGMM is its assumption of Gaussianity. In practice, the intrinsic distribution of TRGB stars within a given magnitude range is far from Gaussian. More complex models of this distribution should be investigated, ideally on high-precision photometry of systems with well-populated RGB sequences. For example, a multicomponent Gaussian mixture could be of use in distinguishing remaining contaminants, such as a low-density AGB “background,” from the TRGB population. Alternate fitting methods, such as Gaussian process regression, should also be considered. We reserve tests of this nature for future exploration using local galaxies with properties and observations that are better matched to the requirements of drawing conclusions from such tests.

### 5.3. Absolute Calibration

In Section 4.2.2, Figures 13 and 14 introduce a puzzle with regard to using synthetic photometry as the absolute calibration for the TRGB. Inspection of Figures 13 and 14 reveals two concerns for the isochrone sets: (i) a systematic offset between PARSEC and MIST is present for the absolute magnitudes regardless of filter, such that PARSEC is consistently brighter than MIST, and (ii) there is a relative filter-to-filter offset between the models’ predictions and our measurements, such that the absolute near-IR magnitudes of our measurements derived by adopting the models’ optical predictions do not correspond to the models’ near-IR predictions, and vice versa. Understanding these differences requires considering how stellar interior models, which predict stellar structure, are mapped to stellar atmospheres used to construct the synthetic photometry shown in Figures 13 and 14. An excellent discussion of this process is given by Casagrande & VandenBerg (2014).

Before examining the models more closely, we note that such offsets should not be surprising when viewed in the context of the larger literature. Even in the well-studied F814W/I band, there is a current debate in the value of the absolute magnitude of the TRGB, which is central to determining $H_0$ (see, e.g., Freedman et al. 2019, 2020; Reid et al. 2019; Yuan et al. 2019, and references therein). Even the most detailed and careful calibrations have total uncertainties at the 0.05 mag level due to various systematic terms. The range of recently used F814W absolute values spans $\sim$0.10 mag (as reviewed by Beaton et al. 2018). These discrepancies in the absolute magnitude of the tip can propagate into stellar models depending on exactly how the isochrone sets cross-check their own absolute scales. As reviewed by Beaton et al. (2018), such discrepancies also affect the RR Lyrae and horizontal branches, which imparts uncertainty on the absolute scale for globular clusters (as are explored in detail by Casagrande & VandenBerg 2014). Therefore, no theoretical predictions can be expected to be immune to the downstream effects of systematics in the empirical distance scale.
In this subsection, we first present a preliminary comparison of our results to the empirical TRGB relation derived by Jang & Lee (2017a) and discuss its implications with regard to assessing differences in the models’ behavior. We then consider two aspects of the synthetic photometry that might contribute to the discrepancies in Figures 13 and 14: (i) differences in adopted stellar atmospheres, which affect the conversion from bolometric luminosity to observed fluxes in specific filters, discussed in Section 5.3.2, and (ii) differences in the underlying stellar evolution physics, discussed in Section 5.3.3.

5.3.1. Comparison to Empirical Optical Results

The discrepancies between our measurements and both sets of synthetic photometry in Figures 13 and 14 make it unclear which model should be preferred, if either. As an alternative, we turn to the empirical F814W TRGB relation presented by Jang & Lee (2017a). Like the majority of existing empirical F814W/IR calibrations, it is based on a specific optical color baseline (F606W–F814W), which precludes us from adopting it for our entire sample, as F814W is the only optical filter common to all our targets. However, the subset of our sample with F606W observations (14 out of 23 targets) allows us to make a preliminary comparison as a benchmark against theoretical calibrations.7

Jang & Lee (2017a) employ a quadratic functional form (the QT system) for the $M_{\text{F814W}}$ versus (F606W–F814W) relation, which they calibrate over an extensive color range of 0.8 mag $<$ F606W $-$ F814W $<$ 3 mag. We show the results of adopting the QT relation for the galaxies in our sample with F606W coverage in Figure 16.

5.3.2. Bolometric Corrections

We make a direct comparison of the models’ physical predictions in the left panel of Figure 17, which compares the MIST and PARSEC model TRGB in temperature–luminosity space for the same range of ages and metallicities as in Figures 13 and 14. We see that the models are offset from each other in $T_{\text{eff}}$ and log($L/L_\odot$), indicating differences in the underlying stellar structure, with PARSEC running $\sim$10% more luminous and 50–150 K warmer than MIST at the same age and metallicity. (For comparison, Choi et al. 2018 find uncertainties on the absolute $T_{\text{eff}}$ scale of $\pm$100 K due to boundary conditions.) The PARSEC predictions also show a slightly larger spread in $T_{\text{eff}}$ than MIST over the same range of age and metallicity.

We attempt to isolate filter-to-filter differences between the models’ predictions by examining their color–color behavior. The middle panel of Figure 17 shows the F814W–F110W to F110W–F160W color–color behavior of the two model sets. There is a divergence on the order of 0.1 mag for stars with F814W–F110W $>$ 1 mag, indicating significant differences in the model atmospheres of cooler stars. Close inspection of the bluer side shows that although the color–color relations have

---

7 A more detailed analysis that incorporates recently revised distances to the two absolute-scale zero-point anchors used by Jang & Lee (2017a) is currently in preparation.
similar slopes, they are slightly offset from one another at the same metallicity.

We compare these color–color predictions to our measurements in the right panel of Figure 17, which overlays the color–color results derived via the MCR-TRGB technique. Unlike Figures 13 and 14, the color–color relations are distance independent, which alleviates concerns about the absolute scale of the measurements. The observed TRGB is systematically redder than predicted in F110W–F160W and/or systematically bluer in F814W–F110W (which appears to effectively rule out unaccounted-for extinction as a source of disagreement). Comparison to the right panel of Figure 16, which shows good agreement between observations and PARSEC predictions for F606W–F814W and F110W–F160W colors, suggests that PARSEC’s predicted TRGB colors are overall accurate in the optical and IR independently, but that there may be offsets in the relative cross-calibration between the two wavelength regimes in either the stellar atmosphere models or our data.

More specifically, if PARSEC’s predicted F110W and F160W absolute magnitudes were shifted to be ~0.1 mag dimmer (or if our near-IR measurements were 0.1 mag brighter) with no changes to the optical, the predicted and observed F814W–F110W colors would be brought into alignment, with no change to the F110W–F160W or F606W–F814W colors.

The stellar atmospheres used to generate synthetic photometry are encoded as bolometric corrections, which transform bolometric luminosities into filter-specific quantities. The bolometric corrections depend primarily on log(g), $T_{\text{eff}}$, and both [Fe/H] and [$\alpha$/H]. While the differences in predicted bolometric luminosities shown in the left panel of Figure 17 suggest that PARSEC and MIST would still predict different absolute magnitudes for a tip star of a given age and metallicity, bolometric corrections and the stellar $T_{\text{eff}}$ scale are likely to be a source of some of the filter-to-filter offsets we observe.
As of this writing, the PARSEC web service (CMD v. 3.3) uses PHOENIX (Allard et al. 2012) bolometric corrections for stars with $T_{\text{eff}} < 5500$ K ($\sim$1000 K warmer than the warmest TRGB stars), and MIST uses ATLAS12 (Kurucz 2014). Chen et al. (2019) have explored how the PHOENIX (Allard et al. 2012) and ATLAS9 (Kurucz 2014) model atmospheres affect predicted colors and magnitudes of PARSEC isochrones in optical and IR passbands. Their Figure 2 shows that the PHOENIX bolometric corrections produce RGB colors that are biased red by up to 0.1 mag in $V - I$, which translates to an artificially bright TRGB in the near-IR consistent with what we see in Figures 13 and 14. Although Chen et al. (2019) claim that the PHOENIX bolometric corrections are preferable for giants because they are computed with spherical geometry, Fu et al. (2018, Section 3.2.3) find that the PHOENIX bolometric corrections cannot reproduce the observed RGB colors in 47 Tuc, which trouble the “RGB-too-red” problem. While we have been unable to locate any similar such studies of MIST’s predictions for the RGB, Fu et al. (2018) caution that ATLAS12 atmospheres may be unreliable for $T_{\text{eff}} < 4000$ K, which may explain the divergence we see between the MIST and PARSEC predictions at F814W–F110W > 1 in the middle panel of Figure 17.

Similar evidence for the importance of bolometric corrections for the optical TRGB was explored by Serenelli et al. (2017), who directly compared the predicted absolute magnitudes of the $I$-band TRGB from the BaSTI models (Pietrinferni et al. 2013) using four sets of bolometric corrections (see their Figure 8). They see differences at the $\sim$10% level when applying different sets of bolometric corrections to models using the same underlying physics, comparable to the amplitude of discrepancies we see here. Although this investigation focused on the optical, their finding of $\sim$0.1 mag discrepancies aligns with the scale of the adjustments needed to bring our measurements and the models’ predictions into alignment.

While further quantitative investigation is clearly required, we conclude that bolometric corrections are a likely source of a substantial part of the optical–IR discrepancies we observe in Figures 13 and 14.

5.3.3. Physical Properties of TRGB Stars

The left panel of Figure 17 suggests that there are currently real differences in the predicted stellar structure at the TRGB owing to the different physical assumptions between the models, even before differences in atmospheres or bolometric corrections are included. Our comparison of PARSEC and MIST broadly agrees with the conclusions from a more detailed model-focused study by Serenelli et al. (2017), who investigated both the physical and computational factors contributing to differences between the predicted TRGB luminosities of two sets of stellar models (BaSTI and GARSTEC). Serenelli et al. (2017) were able to produce identical predictions of tip stars’ physical properties from two different model suites only when certain physical processes, such as neutrino energy loss and electron screening, as well as some numerical criteria such as integration time step, were implemented consistently between stellar evolution codes (the full set of which they term “concordance physics”). While a comparable investigation of such sources of difference between MIST and PARSEC is well outside the scope of this work, we find it reasonable to conclude that some aspects of their physical differences are likely due to limitations in our current understanding of certain “cutting edge” topics in stellar astrophysics and may also be in part due to differing computational approaches. Tip stars, in addition to being both cool and luminous, are at an evolutionary transition point and so may be especially sensitive to these details.

Another possible source of disagreement between the models’ predictions and our data is elemental abundances, including the helium fraction $Y$ and $\alpha$ enhancement, the latter of which is of particular concern at $[\text{Fe}/\text{H}] \lesssim -1$ dex (F110W–F160W $\lesssim 0.9$ mag for PARSEC). At present, neither MIST nor PARSEC has publicly available $\alpha$-enhanced models, although they are slated to be included in future releases of both (Choi et al. 2016; Fu et al. 2018). Serenelli et al. (2017) report that, at least after the adoption of their concordance physics, $\alpha$-enhancement produces what they consider to be negligible effects on the TRGB bolometric luminosity (<1% difference between $[\alpha/\text{Fe}] = 0.4$ and $[\alpha/\text{Fe}] = 0$ at fixed $[\text{M}/\text{H}]$, $T_{\text{eff}}$ (<2% difference), and predicted $V/IJK$ magnitudes (< 0.01 mag difference at constant color). Similarly, they find...
that a change in the helium fraction $Y$ of 0.01 (approximately the range over which estimates of the primordial helium mass fraction vary) has no more than a 1% effect on TRGB temperatures and luminosities. Nonetheless, as forthcoming versions of PARSEC will offer options for variable $\alpha$-enhancement and helium abundance (Fu et al. 2018), as well as updated bolometric corrections,\footnote{The PHOENIX bolometric corrections currently employed in the PARSEC web service only take total $Z$ into account in their color transformations (Fu et al. 2018), which is problematic for understanding potential photometric effects of varying abundance ratios. ATLAS12 model atmospheres for $\alpha$-enhanced PARSEC isochrones down to $T_{\text{eff}} = 4000$ K are currently in development (Fu et al. 2018; Chen et al. 2019).} we anticipate that a direct analysis of these quantities’ impacts on predictions of TRGB behavior as they pertain to this work will be both easily achievable and informative.

5.4. TP-AGB Contamination

While we believe that our method of determining $P$(RGB)$^+$ is overall effective at rejecting contaminating populations, such as red supergiants and the bulk of the AGB, there may be some amount of remaining contamination, particularly from thermally pulsing AGB (TP-AGB) stars, which we briefly discuss here.

Although TP-AGB stars are generally intrinsically brighter than the TRGB, extinction from circumstellar dust can substantially impact their observed magnitudes and bring them closer to luminosities typical of the upper RGB. However, as they are also heavily reddened, their colors are inconsistent with RGB stars (Boyer et al. 2017, see their Figure 8), so it is likely that our method of TRGB candidate selection successfully rejected most, if not all, of these stars.

TiO absorption is another factor that may bring certain TP-AGB stars, particularly M-type stars at high metallicity, closer in luminosity to the TRGB (Boyer et al. 2019). While the bulk of our targets are low-metallicity dwarf galaxies, this may be an issue for some of the larger galaxies in our sample, such as M81.

Finally, TP-AGB stars may cross the TRGB when they reach the minimum point in their pulsation cycle. In this case, the fact that our near-IR data were taken several years later than our optical data is an advantage; TP-AGB stars at their minimum in our optical observations are unlikely to be at their minimum in the near-IR, and vice versa.

5.5. Limited Empirical Constraints on TRGB Magnitude Stability

We briefly discuss other physical concerns that may affect the intrinsic photometric properties of a TRGB star through mass transfer in a binary interaction. Preliminary investigations suggest that the photometric effects of the former phenomenon are overall secondary to the variation of TRGB magnitude with metallicity (J. J. Eldridge 2019, private communication). Thus, we qualitatively conclude that binarity is likely to be a source of some amount of residual scatter in our measurements rather than a primary driver of TRGB variation across populations.

Second, as recent high-precision, high-cadence photometry has demonstrated, low-amplitude variability exists for many to most stellar types. Pulsational variability was first proposed for stars on the upper RGB by Ita et al. (2002) and has since been observationally confirmed (Ita et al. 2004; Lebzelter & Wood 2005; Wood 2015). Variability may contribute some amount of uncertainty to TRGB measurements by effectively blurring the TRGB edge. However, there are few established constraints on relevant characteristics, such as typical periods, amplitudes, fractions of stars that exhibit variability, and dependence on stellar properties such as age and metallicity. We have thus disregarded TRGB variability as a potential systematic in this work owing to lack of empirical constraints. We expect that any overall effects are small compared to our dominant sources of uncertainty.

RGB stars are known to experience mass loss driven by chromospheric activity (Origlia et al. 2007; Groenewegen 2012; Pasquato et al. 2014), which may be amplified by either of the first two properties. Jimenez et al. (2020) predict that variations in the mass-loss parameter $\eta$ at the TRGB may affect individual stars’ luminosities by over 5%, although they estimate that the net effect on measured TRGB distances does not exceed 2% and that it is strongly metallicity dependent. Additionally, although mass loss has been correlated with the blueshifting of optical and near-IR spectral lines such as H$\alpha$ and the calcium...
triplet (McDonald & van Loon 2007; Wood 2015), the impact of this blueshifting on broadband photometry has not been quantified.

Jimenez et al. (2020) also consider planetary engulfment, wherein an RGB star consumes one or more planets in close orbit as it expands. They predict that the increased turbulence in the star’s convective envelope, corresponding to an increase in mixing length, may result in a net decrease in TRGB luminosity by up to 5% for a single star that has consumed a giant planet. However, they conclude that both detailed hydrodynamical simulations and further studies of planetary system formation are required to accurately constrain potential impacts of this phenomenon on the TRGB as a distance indicator.

Again, we expect that these effects are overall well within the uncertainties of this work, but they may need to be taken under consideration in future high-precision TRGB studies.

6. Conclusions and Future Work

6.1. Conclusions

We have developed a method to measure TRGB magnitudes and colors in multiple filters simultaneously. This method, MCR-TRGB, was designed to use a set of likely RGB stars, which were defined where traditional TRGB detection methods using edge detection can be employed reliably, to study the multiwavelength behavior of the TRGB using those same stars. We applied MCR-TRGB to a re-reduction of optical + near-IR HST data originally presented in D12; these new reductions use the optical observations, which have higher spatial resolution and are generally more complete at the TRGB, to produce more complete and precise photometry in the infrared bands. When using the same distances as D12, we find only minor adjustments to the color–magnitude behavior of the IR-TRGB. However, the D12 absolute magnitudes were determined relative to color–magnitude predictions from stellar models. Thus, we compared three different absolute magnitude calibrations of the measured TRGB magnitudes, one using the same distance moduli as in D12, and two using distance moduli derived from the predicted TRGB absolute magnitudes from two commonly used isochrone sets (PARSEC and MIST). We find that the isochrone-based absolute calibrations are inconsistent with each other at the ∼0.1 mag level, consistent with previous work in this domain, and that both sets of isochrones are internally inconsistent with our measurements of the TRGB magnitudes and colors when optical and infrared measurements are used together. We further caution that when adopting model-based absolute calibrations for the TRGB, a conservative 10% systematic should be included in the calibrations for the TRGB based on differences between the isochrone sets at the TRGB. We find that these tensions persist even with the application of a state-of-the-art empirical calibration. From examining the distance-independent color–color behavior of our data against model predictions, we conclude that bolometric corrections and the underlying stellar $T_{\text{eff}}$ scale are likely to explain a large part of the inconsistencies we have found.

6.2. Future Work

An empirical absolute TRGB calibration in WFC3/IR bandpasses remains elusive. A fully model-independent calibration, as in Jang & Lee (2017a), is clearly necessary. However, there are a limited number of systems that are distant enough that their apparent TRGB magnitudes are easily observable with HST but nearby enough to have distances that are well constrained by other means.

Another limitation of our study is the lack of precise independent distances to the galaxies in this work. Indeed, the majority of these systems only have distances determined from the TRGB itself. Some of these galaxies are within a volume for variable-star-based distances with HST, though we are cautious about their precision given the metallicity dependence of such relations and the difficulty of inferring stellar metallicities for galaxies at these distances (for RR Lyrae see Beaton et al. 2018).

Our initial goal in this work was to relate empirical results on the multiwavelength TRGB to the physical characteristics of the underlying stellar populations, such as age and metallicity. We found that goal challenging owing to the internal mismatches we observe in the isochrone sets, given that the aforementioned physical parameters are ultimately inferred via comparison to those from isochrone sets once a distance, also often isochrone dependent, is assumed.

We gratefully acknowledge the anonymous referee, Emily Levesque, Željko Ivezić, Anil Seth, and Evan Skillman for helpful feedback on this manuscript, as well as J. J. Eldridge, Olivia Jones, and several members of the Carnegie-Chicago Hubble Program team for illuminating discussions. We also thank the GalRead group at Princeton University, in particular Andy Goulding, for directing us to Extreme Deconvolution.

We acknowledge the people of the Dkhw’Duw’Absh, the Duwamish Tribe, the Muckleshoot Tribe, the Lenape, and other tribes on whose traditional lands we have performed this work.

Support for this work was provided by NASA through grant No. AR-15016 and through Hubble Fellowship grant No. 51386.01 awarded to R.L.B., from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS 5–26555.

This research has made use of “Aladin sky atlas” developed at CDS, Strasbourg Observatory, France, and of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Facilities: HST(ACS/WFC), HST(WFC3/IR).

Software: AstroML (VanderPlas et al. 2012, 2014), Astropy (Astropy Collaboration et al. 2013, 2018), Astroquery (Ginsburg et al. 2017, 2019), Dask (Rocklin 2015; Dask Development Team 2016), DOLPHOT (Dolphin 2000, 2016), Drizzlepac (STSCI Development Team 2012; Hack et al. 2013; Avila et al. 2015), KDEpy (Odlnd 2018), Matplotlib (Hunter 2007), NumPy (van der Walt et al. 2011), Pandas (Mckinney 2010, 2011), Seaborn (Waskom et al. 2018), SciPy (Jones et al. 2001), SciKit-learn (Pedregosa et al. 2011), SEP (Barbary 2016, 2018), Vax (Breddels & Veljanoski 2018a, 2018b).

Appendix

Tests on Artificial Data

In this appendix we diagnose potential biases and systematics induced by our technique by applying the above methods to simulated CMDs with known theoretical TRGB magnitudes. We use the results of this analysis to determine bias corrections
to our final tip magnitudes and to refine the uncertainties on our measurements.

We first generate a set of idealized (i.e., error-free) photometry of artificial RGB sequences with MATCH based on the PARSEC model suite. Serenelli et al. (2017) demonstrated that metallicity is the primary driver of variation in TRGB colors and magnitudes for old ages, so we hold all parameters except metallicity constant. We use a Chabrier IMF with a slope of 1.3, a binary fraction of 0.3, and a constant star formation rate (SFR) with an age range of 100 Myr–14 Gyr. We vary metallicity between $-2.0$ dex $<$ [Fe/H] $< -0.5$ dex with a spacing of 0.1 dex, which covers a color range representative of our data.

We constrain the output magnitudes to $\sim 1$ mag brighter than the TRGB to $\sim 2.5$ mag dimmer for each filter: $-5.0$ mag $<$ F814W $< -1.5$ mag, $-6.0$ mag $<$ F110W $< -3.0$ mag, and $-7.0$ mag $<$ F160W $< -3.5$ mag.

We choose to model a constant SFR rather than a single-age population for two reasons. First, most of our data (see star formation histories in D12) show evidence for stellar sequences from young populations, and as such, mono-age populations are not realistic representations of our data. Second, for a mono-age population, the full color or magnitude range of the TRGB is small enough to be comparable to the typical errors on the data points. In this limit, XDGMM cannot resolve the underlying distribution of TRGB color and magnitude, and thus its performance cannot be tested.

We add fiducial photometric errors to the simulated stellar population using the aggregate of all of our AST results, which were determined for each galaxy in our data set in Section 2.3. We do not incorporate a net photometric bias or photometric incompleteness, both of which are negligible at the TRGB for the majority of our sample.

For all tests, we randomly subsample the artificial data to reach a desired total number of RGB stars less than 1 mag fainter than the fiducial TRGB ($N_{T+1}^{\pm}$) in the filter used to measure the Sobel edge. We also adjust all magnitudes by a value randomly generated from a Gaussian with a mean of 0 mag and standard deviation set to the star’s photometric uncertainty.

### A.1. LF Sampling

Here we test our method against $N_{T+1}^{\pm}$ over a range of $200 \leq N_{T+1}^{\pm} \leq 5000$ stars, which spans the $N_{T+1}^{\pm}$ values for the majority of the galaxies in our sample. At each $N_{T+1}^{\pm}$ we run 20 end-to-end XDGMM tip-fitting iterations and calculate the offsets $\Delta M_{TRGB} \equiv M_{TRGB}(\text{measured}) - M_{TRGB}(\text{true})$ for each filter. We repeat these tests at four different metallicities ([Fe/H] = [−1.7, −1.3, −1.0, −0.7] dex) to check for possible color-dependent effects.

Figure A1 shows the median per-filter differences in the measured versus theoretical TRGB values against $N_{T+1}^{\pm}$ for each of the four metallicities. The error bars on the circles in Figure A1 show the interquartile range of the results. The dashed horizontal lines are color-coded to match the filter and show the range of the per-filter mean uncertainty on the tip-fitting results; we define the uncertainty as the quadrature sum of the XDGMM fitting uncertainty and the median photometric error of the tip stars.

For all but [Fe/H] = −0.7 dex, the results are largely consistent: the offsets $\Delta M_{TRGB}$ start out around 0.04 mag in the most undersampled case, increase approximately with $N_{T+1}^{\pm}$ for $N_{T+1}^{\pm} \sim 1000$ stars in all filters, and then begin to level off near $\Delta M_{TRGB} \sim 0.01$ mag. These results are broadly similar to what is seen for traditional edge detection methods. Madore et al. (2009) found that Sobel edge detection is prone to bias when the RGB LF is undersampled and that a sample of $N_{T+1}^{\pm} \gtrsim 500$ stars is required for edge detection to function accurately.

For [Fe/H] = −0.7 dex, where the edge detection and initial tip star selection are done in F110W rather than F814W, we see behavior similar to the lower metallicities in the near-IR, but not in F814W, which hovers at 0.04 mag throughout. This foreshadows a possible selection effect when using F110W for the primary edge detection, which we investigate further in Section A.2.

### A.2. Metallicity

For each metallicity in our artificial data set, we run 20 end-to-end tip-fitting iterations with $N_{T+1}^{\pm} = 4000$ stars, in the regime where sampling effects are minimal. Figure A2 shows the median per-filter differences in the measured versus the theoretical TRGB values against the measured IR-TRGB color; the IR-TRGB color increases approximately monotonically with metallicity in the artificial data.

The jump in offset values at F110W–F160W $> 0.95$ mag corresponds to the switch from using F814W to using F110W for the Sobel edge detection. We hypothesize that this jump is due to a difference in which stars are selected as candidate tip stars. At high metallicity, stars that have similar magnitudes in F110W may have a large range of magnitudes in F814W owing to the increasingly steep TRGB color slope in F814W. The stars that are the brightest in F814W are relatively dim in F110W and so may not fall within the F110W tip star selection window, and thus our measured F814W tip magnitudes are skewed faint relative to the predicted values.

### A.3. Photometric Errors

To isolate the impact of photometric uncertainties on our TRGB measurements, we use the artificial data set with $N_{T+1}^{\pm} = 2000$ stars and [Fe/H] = −1.0 dex. For each of 500 trials, we vary the input magnitudes by a different random value drawn from a Gaussian whose standard deviation is each star’s photometric uncertainty. The results of this test are shown in Figure A3. We find that for all filters the standard deviation of the resulting distribution of offsets is roughly a third of the typical photometric uncertainty at the tip.

### A.4. XDGMM versus Sobel Edge Detection

Here we investigate the behavior of XDGMM tip fitting relative to the standard method of Sobel edge detection using the same set of trials as in Section A.1. As our method for XDGMM tip fitting itself uses Sobel edge detection to set the color–magnitude center of the initial tip star selection window, we can make a fully self-consistent comparison of the Sobel edge magnitudes to the XDGMM mean magnitudes for detections in F814W and F110W. (We do not perform edge detection on F160W in our method and so do not make the comparison.)

The top panel of Figure A4 shows the relation between the median difference between the XDGMM-fitted mean and the theoretical tip versus the median difference between the Sobel edge and theoretical tip. The linear fit to the data is consistent
within 1σ with a one-to-one relation, indicating that the methods produce overall consistent tip magnitudes.

The bottom panel of Figure A4 shows the relation between the width of the tip star selection region, Δη, and the difference between the XDGMM- and Sobel-derived tip magnitudes, ΔMTRGB. The quantities are correlated, albeit with some scatter on the order of 0.01 mag, and are fit by the linear relation $\Delta M_{\text{TRGB}} = 0.25(\Delta \eta) - 0.01$ mag.

### A.5. Adjustments to Measurements

In the previous subsections, a number of tests were performed to quantify the statistical and systematic uncertainties of the MCR method using artificial photometry. Here we match our observed galaxies to their artificial tests to determine both systematic terms that are applied in the form of bias corrections and statistical terms that are applied in the form of inflating the algorithmic uncertainties. The corrections are parameterized by two key observables: (i) how well populated the RGB is as a proxy for the total mass, and (ii) the F814W–F160W color as a proxy for the underlying stellar population properties. All such adjustments are summarized in Table A1, and if a given target does not appear in the table, then it did not require a modification.

For each target, we determine the most appropriate sets of tests to use to determine the bias based on $N_*^{T^+1}$ and F814W–F160W color. The quoted adjustment values are adopted from the relevant set of trials, with uncertainties determined as the median and interquartile range of the offsets (measured − predicted value) in each filter. We subtract the offsets from the measured TRGB apparent magnitude and add the associated uncertainty in quadrature to the fitting uncertainty. We also modify all relevant colors based on these adjustments.

The first four targets in Table A1 (KDG 73, NGC 2403-HALO-6, SCL-DE1, and UGCA 292) all have $N_*^{T^+1} < 500$.
Although these targets do not all have the same colors, we found that differences between offsets were negligible at the relevant colors. For these we take the median and interquartile range of offsets for all trials with \( N = 500 \) stars and \([\text{Fe}/\text{H}] = -1.0 \) dex. The remaining targets (NGC 0300, NGC 2976-DEEP, NGC 3077-PHOENIX, SN-NGC 2403-PR, and M81-DEEP) use F110W as the edge detection filter. All but M81-DEEP have colors F814W − F160W ∼ 2 mag, whereas M81-DEEP has F814W − F160W ∼ 2.25 mag. We match these colors by adopting the median and interquartile range of offsets for trials with −0.8 dex ≤ [Fe/H] ≤ −0.7 dex for all but M81-DEEP and −0.6 dex ≤ [Fe/H] ≤ −0.5 dex for M81-DEEP.

Table A1

| Target Name   | Value±Error Subtracted from \( m_{\text{TRGB}} \) |
|---------------|--------------------------------------------------|
| KDG 73        | 0.04 ± 0.02                                      |
| NGC 2403-HALO-6 | 0.04 ± 0.02                                     |
| SCL-DE1       | 0.04 ± 0.02                                      |
| UGCA 292      | 0.04 ± 0.02                                      |
| NGC 0300      | 0.03 ± 0.01                                      |
| NGC 2976-DEEP | 0.03 ± 0.01                                      |
| NGC 3077-PHOENIX | 0.03 ± 0.01                                    |
| SN-NGC 2403-PR | 0.03 ± 0.01                                     |
| M81-DEEP      | 0.06 ± 0.01                                      |

Figure A2. Difference between measured and theoretical TRGB values against the median measured near-IR tip color for the ensemble of simulated data sets. Results for F814W, F110W, and F160W are shown in blue, orange, and green, respectively. The horizontal lines show ±1σ, where σ is the mean quadrature sum of the photometric and fitting errors for each filter. This bias is within the TRGB detection uncertainty in almost all cases.

Figure A3. Histograms of the difference between the measured and predicted \( m_{\text{TRGB}} \) for the artificial data set with \( N^+ = 2000 \) stars and \([\text{Fe}/\text{H}] = -1.0 \) dex for 500 trials, where each trial modifies the stellar magnitudes randomly in proportion to their photometric errors. The blue histogram is for F814W, the orange is for F110W, and the green is for F160W. As expected, the widths of all histograms are smaller than the reported photometric uncertainties at the tip.

Figure A4. Top: median difference between XDGMM-fitted and theoretical tip magnitudes vs. the median difference between the Sobel edge magnitude and theoretical tip magnitude in either F814W or F110W. Each point is the median result for one set of trials with fixed metallicity and \( N^+ = 1 \). The solid line and flanking filled region show a linear fit to the data and its 68% confidence interval, whereas the dashed line shows a one-to-one relation. Bottom: difference between the XDGMM-fitted mean and Sobel edge magnitude vs. \( \Delta \eta \), which is the width in magnitudes of the tip star selection region in the LF. Each point represents the offset for a single trial with fixed metallicity and \( N^+ = 1 \). The solid line and flanking filled region show a linear fit to the data and its a 95% confidence interval, respectively.

The Astrophysical Journal, 898:57 (26pp), 2020 July 20 Durbin et al.
STSCI Development Team 2012, DrizzlePac: HST Image Software, ascl:1212.011
Swam, M. S., Lubow, S., & Hurt, L. 2004, in ASP Conf. Ser. 314, Astronomical Data Analysis Software and Systems (ADASS) XIII, ed. F. Ochsenbein, M. G. Allen, & D. Egret (San Francisco, CA: ASP), 824
Sweigart, A. V., & Gross, P. G. 1978, ApJS, 36, 405
Trujillo, I., Beasley, M. A., Borlaff, A., et al. 2019, MNRAS, 486, 1192
Tully, R. B., Courtois, H. M., Dolphin, A. E., et al. 2013, AJ, 146, 86
Tully, R. B., Rizzi, L., Dolphin, A. E., et al. 2006, AJ, 132, 729
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
VandenBerg, D. A., Swenson, F. J., Rogers, F. J., Iglesias, C. A., & Alexander, D. R. 2000, ApJ, 532, 430
VanderPlas, J., Connolly, A. J., Ivezic, Z., & Gray, A. 2012, in Proc. of Conf. on Intelligent Data Understanding (CIDU) (Boulder, CO: IEEE), 47
VanderPlas, J., Fouesneau, M., & Taylor, J. 2014, AstroML: Machine Learning and Data Mining in Astronomy v0.4.0, Astrophysics Source Code Library, ascl:1407.018
Waskom, M., Botvinnik, O., O’Kane, D., et al. 2018, Mwaskom/Seaborn: V0.9.0 (July 2018), Zenodo, doi:10.5281/ZENODO.1313201
Williams, B. F., Lang, D., Dalcanton, J. J., et al. 2014, ApJS, 215, 9
Wood, P. R. 2015, MNRAS, 448, 3829
Wu, P.-F., Tully, R. B., Rizzi, L., et al. 2014, AJ, 148, 7
Yuan, W., Riess, A. G., Macri, L. M., Casertano, S., & Scolnic, D. M. 2019, ApJ, 886, 61