The Astrophysical Journal, 956:51 (28pp), 2023 October 10 https://doi.org/10.3847/1538-4357/acefd4


# The RESOLVE and ECO Gas in Galaxy Groups Initiative: The Group Finder and the Group H I–Halo Mass Relation

Zackary L. Hutchens[1], Sheila J. Kannappan[1], Andreas A. Berlind[2,3], Mehnaaz Asad[3], Kathleen D. Eckert[1], David V. Stark[4], Derrick S. Carr[1], Ella R. Castelloe[1], Andrew J. Baker[5,6], Kelley M. Hess[7,8], Amanda J. Moffett[9], Mark A. Norris[10], and Darren Croton[11,12]

[1] Department of Physics and Astronomy, University of North Carolina at Chapel Hill, 120 E. Cameron Ave., Chapel Hill, NC 27599, USA; zhutchen@live.unc.edu
[2] Division of Astronomical Sciences, National Science Foundation, Alexandria, VA 22314, USA
[3] Department of Physics and Astronomy, Vanderbilt University, Nashville, TN 37240, USA
[4] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
[5] Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[6] Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Rd., Bellville 7535, South Africa
[7] Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18008 Granada, Spain
[8] ASTRON, The Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA, Dwingeloo, The Netherlands
[9] Department of Physics and Astronomy, University of North Georgia, 3820 Mundy Mill Rd., Oakwood GA 30566, USA
[10] Jeremiah Horrocks Institute, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK
[11] Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[12] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Melbourne, Australia
*Received 2023 February 20; revised 2023 August 7; accepted 2023 August 8; published 2023 October 11*

## Abstract

We present a four-step group-finding algorithm for the Gas in Galaxy Groups (G3) initiative, a spin-off of the $z \sim 0$ REsolved Spectroscopy Of a Local VolumE (RESOLVE) and Environmental COntext (ECO) surveys. In preparation for future comparisons to intermediate redshift (e.g., the LADUMA survey), we design the group finder to adapt to incomplete, shallow, or nonuniform data. We use mock catalogs to optimize the group finder's performance. Compared to friends-of-friends (with false-pair splitting), the G3 algorithm offers improved completeness and halo-mass recovery with minimal loss of purity. Combining it with the volume-limited H I census data for RESOLVE and ECO, we examine the H I content of galaxy groups as a function of group halo mass. Group-integrated H I mass $M_{\rm H\,I,grp}$ rises monotonically over halo masses $M_{\rm halo} \sim 10^{11}$–$10^{14.5}\,M_{\odot}$, pivoting in slope at $M_{\rm halo} \sim 10^{11.4} M_{\odot}$, the gas-richness threshold scale. We present the first measurement of the scatter in this relation, which has a median of ∼0.3 dex and is asymmetric toward lower $M_{\rm H\,I,grp}$. We discuss interesting tensions with theoretical predictions and prior measurements of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation. In an appendix, we release RESOLVE DR4 and ECO DR3, including updates to survey redshifts, photometry, and group catalogs, as well as a major expansion of the ECO H I inventory with value-added data products.

*Unified Astronomy Thesaurus concepts:* Galaxies (573); Galaxy evolution (594); Galaxy groups (597); Galaxy dark matter halos (1880)

*Supporting material:* machine-readable tables

## 1. Introduction

Neutral atomic hydrogen (H I) is a transitional gas phase that connects the warm, ionized intergalactic medium to the cold, dense gas that galaxies consume for star formation and AGN activity. Thus, many insights into galaxy evolution and cosmology can be attained by quantifying the evolution and distribution of H I in the Universe, and in particular, the group H I–halo mass relation (here, the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation). This scaling relation quantifies the H I mass held in galaxy groups as a function of group dark matter halo mass, including halos containing only one galaxy ($N_{\rm galaxies} = 1$ "groups"). As noted by Obuljen et al. (2019), the relation reflects tremendous astrophysical information, such as cosmic gas accretion (e.g., as in Kereš et al. 2005), stellar and AGN feedback (Di Matteo et al. 2005; Chauhan et al. 2020), or group processes such as starvation and stripping (as in Abadi et al. 1999; Bekki et al. 2002). In cosmology, measurements of the group H I–halo mass relation can constrain the cosmic H I abundance and the clustering of H I, which traces the underlying matter distribution (Castorina & Villaescusa-Navarro 2017). Thus, in an era of deep H I surveys such as LADUMA (Looking At the Distant Universe with the MeerKAT Array; Blyth et al. 2016), there is substantial motivation to measure the $z \sim 0$ group H I–halo mass relation and its cosmic evolution.

At $z \sim 0$, a successful approach to examining relationships between H I and environment has been to combine optical group catalogs and H I surveys (e.g., Hess & Wilcots 2013; Stark et al. 2016; Eckert et al. 2017; Guo et al. 2020), ideally over a wide range of group halo masses and environments. Obuljen et al. (2019) constrained the mean $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation through simultaneous empirical modeling of the clustering of ALFALFA H I sources and the abundance of H I in cross-matched SDSS groups; however, direct measurements of group-integrated H I from ALFALFA were limited to massive halos ($>10^{12.5}\,M_{\odot}$), thus missing direct observations in the interesting regime of group physics ($M_{\rm halo} \sim 10^{11.5}\,M_{\odot}$ to $10^{12.1}\,M_{\odot}$) in which groups transition from gas- to stellar-dominated (Eckert et al. 2017) and in which halos are expected to transition from cold to hot accretion modes (Kereš et al. 2005; Dekel & Birnboim 2006;

Kannappan et al. 2013). Guo et al. (2020) later made a direct observation of the $z \sim 0$ mean $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation by stacking ALFALFA spectra of optically selected groups, covering a vast range in group halo mass, but the downside to stacking is that it conceals the relation's scatter, and both shape and scatter are needed to discriminate theoretical models (e.g., see Chauhan et al. 2020). Thus, we are faced with several gaps in our knowledge of the $z \sim 0 M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation: Does the shape of the relation, or our interpretation thereof, change if we forgo stacking, and if so, why? What is the typical scatter in the relation? Which models best replicate the observed shape and scatter? We must answer these questions to better understand the $z \sim 0$ $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation, which will serve as the baseline for future studies of its cosmic evolution.

At present, our knowledge of the high-$z$ ($z \lesssim 1$) $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation is constrained only by intensity mapping (Spina et al. 2021), which can be subject to foreground contamination from low-redshift H I sources and Galactic radio emission. However, deep H I surveys such as LADUMA will soon make it possible to directly measure the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation by combining high-$z$ optical group catalogs and H I surveys. It is thus crucial that groups be defined consistently at $z \sim 0$ and higher $z$, to enable a fair comparison across redshift. Although numerous group-finding techniques have been designed and applied at low $z$, such as friends-of-friends (FoF; Huchra & Geller 1982) and iterative halo-based algorithms (Yang et al. 2005; Lim et al. 2017), these techniques do not address the likelihood of high-$z$ surveys such as LADUMA not having the same highly complete, uniform, or volume-limited data frequently available for $z \sim 0$ group finding, which poses several issues. As noted by Duarte & Mamon (2014), the application of conventional group-finding algorithms such as FoF to flux-limited data is inherently problematic due to the fact that resulting group metrics vary with the survey absolute magnitude limit as a function of redshift. Moreover, as high-$z$ data sets are likely to be both shallower and less uniform in detection of dwarf galaxies, dwarfs should not be used as the primary basis for defining groups, and conventional algorithms like FoF do not account for this bias. Finally, due to any incomplete, nonuniform, or flux-limited nature of high-$z$ data, halo quantities (e.g., virial radii) cannot be computed reliably using standard techniques such as abundance matching. Therefore, these halo quantities cannot be used for group membership refinement, a practice adopted in many group-finding algorithms (Yang et al. 2005; Eckert et al. 2016; Tempel et al. 2016; Lim et al. 2017).

In this paper, we apply $z \sim 0$ data sets to prepare for future evolutionary measurements of group H I relations. To address group-finding challenges, we present a new group-finding algorithm and corresponding group catalogs for the Gas in Galaxy Groups (G3) initiative, an extension of the ongoing $z \sim 0$ RESOLVE (REsolved Spectroscopy Of a Local VolumE; Kannappan & Wei 2008) and ECO (Environmental COntext; Moffett et al. 2015) surveys.[13] RESOLVE and ECO are volume-limited and exceptionally complete, but our algorithm has been designed for translation to other surveys that may be incomplete, nonuniform, or flux-limited, especially LADUMA. Additionally, we present a major update to ECO's archival H I census by cross-matching ALFALFA-100 sources (Haynes et al. 2018) and computing upper limits and confusion flags, complementing RESOLVE's deep and uniquely comprehensive, albeit smaller, H I census. By combining these H I data with our new G3 group catalogs, we can quantify the shape and scatter of the $z \sim 0$ $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation by summing H I mass in grouped galaxies (as in Obuljen et al. 2019), allowing us to measure the relation without stacking over a diverse range of halo masses and environments. The resulting measurement helps to validate the utility of our new group finder and is complementary to previous work on the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation.

This paper proceeds as follows. In Section 2, we describe the RESOLVE and ECO surveys, and we describe the mock catalogs used to optimize and test our new group-finding algorithm. In Section 3, we describe the mathematical and algorithmic details of our new group finder. In Section 4, we employ mock catalogs to optimize the algorithm and quantify its performance. In Section 5, we measure and discuss the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation realized by our new group finder, and we compare it to previous work. Finally, we summarize our conclusions in Section 6. Appendix gives a detailed description of ECO DR3 and RESOLVE DR4, including updates to the H I inventory.

## 2. Data

### 2.1. RESOLVE

The REsolved Spectroscopy Of a Local VolumE (RESOLVE; Kannappan & Wei 2008) survey is a volume-limited census of stellar, gas, and dynamical mass in 53,000 Mpc$^3$ of the nearby cosmic web. RESOLVE contains >1400 galaxies in two ("A" and "B") equatorial strips. RESOLVE-A, covering $8.75^h \leqslant$ R. A. $\leqslant 15.75^h$ and $0° \leqslant$ decl. $\leqslant 5°$, is complete to an absolute $r$-band magnitude of $M_r = -17.33$. RESOLVE-B, covering $22^h \leqslant$ R. A. $\leqslant 3^h$ and $-1.^\circ25 \leqslant$ decl. $\leqslant +1.^\circ25$, is complete to $M_r = -17.0$. Due to the tight correlation between $r$-band luminosity and baryonic mass (cold gas + stars), RESOLVE is complete in baryonic mass to $M_{\rm bary} \sim 10^{9.3} M_\odot$ and $M_{\rm bary} \sim 10^{9.1} M_\odot$ in the A and B semesters (Kannappan et al. 2013; Eckert et al. 2016). The stellar mass completeness limits for RESOLVE-A and RESOLVE-B are $M_* \sim 10^{8.9} M_\odot$ and $M_* \sim 10^{8.7} M_\odot$, respectively (Kannappan et al. 2013; Eckert et al. 2016).

Membership in the RESOLVE survey is confined to $4500 \leqslant v_{\rm grp}\,[{\rm km\,s^{-1}}] \leqslant 7000$, where $v_{\rm grp}$ is the Local Group-corrected recessional velocity of the galaxy's group[14] from the original group catalog of Eckert et al. (2016). We discuss the pre-existing group catalog and this selection strategy in Section 2.4. RESOLVE contains a uniquely comprehensive, fractional-mass-limited 21 cm census ($1.4 M_{\rm H\,I}/M_* \lesssim 0.05$–$0.1$) constructed from public ALFALFA data and deeper Arecibo and Green Bank observations (Stark et al. 2016), detailed in Section 2.3. RESOLVE also includes high-quality reprocessed multiwavelength photometry (covering UV to NIR) and stellar masses derived from spectral energy distribution modeling (Eckert et al. 2015, hereafter E15). Given its unique H I census and highly complete, volume-limited design, RESOLVE is an ideal laboratory for studying relationships between gas and environment.

With this work, we release RESOLVE DR4 with minor updates to H I parameters and catalog membership, an updated FoF group catalog, and our new G3 groups (Appendix). When performing galaxy group finding in RESOLVE DR4 in Section 3, we enforce a limit of $M_r \leqslant -17.33$ in RESOLVE-A and $M_r \leqslant -17.0$ in RESOLVE-B, yielding a total sample of 1690 galaxies split

[13] http://resolve.astro.unc.edu/

[14] For isolated galaxies ($N = 1$ groups), the group velocity is equivalent to the galaxy velocity.

over the RESOLVE-A ($1.5876 \times 10^4\ h^{-3}$ Mpc$^3$) and RESOLVE-B ($5.677 \times 10^3\ h^{-3}$ Mpc$^3$) volumes.[15] Giant galaxies, which we define as having $M_r < -19.5$, represent 577 of these galaxies. When analyzing the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation in Section 5, we further limit our sample to RESOLVE DR4 G3 groups whose centers lie within $4500 \leqslant v_{\rm grp}\,[{\rm km\,s^{-1}}] \leqslant 7000$. The final sample for this analysis includes 1457 galaxies.

### 2.2. ECO

The Environmental COntext (ECO; Moffett et al. 2015; Eckert et al. 2016) survey is a larger catalog surrounding RESOLVE-A in a 440, $000 Mpc^3$ volume. ECO contains approximately 12, 600 galaxies in the volume defined by $8.6^h \leqslant$ R. A. $\leqslant 15.8^h$, $-1^\circ \leqslant$ decl. $\leqslant +50^\circ$, and $3000 \leqslant v_{\rm grp}\,[{\rm km\,s^{-1}}] \leqslant 7000$. ECO reaches the same completeness limits as RESOLVE-A ($M_r = -17.33$). Compared to RESOLVE, ECO is purely archival and includes no new photometry, optical spectroscopy, or 21 cm observations, except where new observations have been inherited in the overlap with RESOLVE-A. All archival data sets used to construct ECO, such as SDSS and ALFALFA, are processed with RESOLVE's pipelines to harmonize the two catalogs. Although ECO's H I data are shallow and flux-limited compared to RESOLVE, its order-of-magnitude larger volume enables robust studies of galaxy environment, including cosmic variance calibrations for the higher-quality RESOLVE survey.

With this work, we release ECO DR3, which includes updates to survey membership, galaxy redshifts, and the FoF group catalog, as well as our new G3 groups and a major addition of archival 21 cm data (Appendix). When performing galaxy group finding in Section 3, we limit ECO DR3 to galaxies for which $M_r \leqslant -17.33$ and $2530 \leqslant v_{\rm grp}\,[{\rm km\,s^{-1}}] \leqslant 7470$, resulting in a sample of 12,771 galaxies in our updated volume estimate[3] of $1.91936 \times 10^5\ h^{-3}{\rm Mpc}^3$. Of these, 4,714 are giants. Our analysis of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation in Section 5 is additionally limited to ECO DR3 G3 groups for which $3000 \leqslant v_{\rm grp}\,[{\rm km\,s^{-1}}] \leqslant 7000$, yielding a sample of 9640 galaxies. Both group finding and group H I analysis exclude ECO database entries flagged as duplicates (see Section A.2).

### 2.3. 21 cm Data

This work leverages the highly complete 21 cm data contained within RESOLVE and ECO to examine the H I mass of galaxy groups. The RESOLVE atomic gas census (Stark et al. 2016, hereafter S16) provides a 21 cm detection or strong upper limit ($M_{\rm H\,I} \lesssim 0.05$ to $0.1 M_*$) for ∼94% of galaxies in the survey. The census combines original, deep observations from Arecibo and the Green Bank Telescope with archival data from ALFALFA (Haynes et al. 2011) and Springob et al. (2005). Detections were measured using the location of the optical galaxy as a prior, reporting $3\sigma$ upper limits for any galaxies that were not detected. Upper limits were computed using the $r$-band Tully–Fisher relation of Kannappan et al. (2013) to estimate 21 cm profile linewidths. Confusion flags were assigned to galaxies by automatically searching for companions in existing redshift surveys. The 21 cm spectra for all flagged galaxies were inspected manually, and deconfusion was attempted as described in S16.

We also use 21 cm data from the ECO Data Release 3 (Appendix), which includes 21 cm detections or upper limits for ∼80% of ECO galaxies. It is comprised of two overlapping parent data sets. Because ECO completely surrounds RESOLVE-A, all A-semester data from the RESOLVE 21 cm census were inherited by ECO, including $3\sigma$ upper limits, confusion flags, and deconfused detections. For the rest of ECO and as a second data option in the RESOLVE-A region (enabling uniform ECO H I data if desired), 21 cm data were obtained from ALFALFA-100 ($\alpha$100; Haynes et al. 2018). Confusion flags and upper limits were computed for ECO galaxies in the $\alpha$100 footprint following the methodology of S16. For ECO, H I upper limits were computed at the $5\sigma$ level to match the reporting threshold of blind 21 cm detections in the published ALFALFA catalog. Unlike for the original RESOLVE data set, no deconfusion is attempted for the secondary, larger ECO data set. In this work, we opt to use the nonuniform mixture of high-quality, fractional-mass-limited data in RESOLVE-A and $\alpha$100-based data in the rest of ECO.

For both RESOLVE and ECO, we compute a single "best" $\log M_{\rm H\,I}$ estimate for each galaxy using detection, upper limit, confusion, and photometric gas fraction information, as further described in Appendix.

### 2.4. FoF Group Catalogs

Eckert et al. (2017), hereafter E17, provide friends-of-friends group catalogs for the RESOLVE and ECO surveys using the algorithm described by Berlind et al. (2006) with linking lengths of $b_{\parallel} = 1.1$ and $b_{\perp} = 0.07$ (see Eckert et al. 2016, hereafter E16, and Duarte & Mamon 2014). After FoF group finding, E17 applied an algorithm to break false galaxy pairs. ECO galaxies are selected for group finding if they meet the completeness limit ($M_r \leqslant -17.33$) and have redshifts satisfying $2530 \leqslant cz\,[{\rm km\,s^{-1}}] \leqslant 7470$. This provides a $\pm 470$ km s$^{-1}$ buffer around the official ECO survey volume. The buffer region mitigates group-finding errors incurred by applying FoF at the edges of the survey. Galaxies are only included in the official ECO survey if the group redshifts lie within $3000 \leqslant v_{\rm grp}\,[{\rm km\,s^{-1}}] \leqslant 7000$, so very few groups are unintentionally clipped by the edges of the survey (the most notable is Coma; see Moffett et al. 2015).

Group memberships in RESOLVE-A were extracted directly from the ECO group catalog, because ECO extends around RESOLVE-A with a 1 Mpc spatial buffer on all sides. To identify groups in RESOLVE-B, E17 used a "RESOLVE-B analog" data set constructed from ECO, with luminosity floor dropped to $M_r = -17$, to determine the physical linking lengths for use in RESOLVE-B, which is too small and subject to cosmic variance for computing linking lengths directly. FoF was applied to galaxies for which $M_r \leqslant -17.0$ and $4250 \leqslant cz\,[{\rm km\,s^{-1}}] \leqslant 7250$, creating a $\pm 250$ km s$^{-1}$ buffer around the official B-semester volume, which again mitigates edge effects and is used to determine survey membership.

We note that M15 and E16 used two completeness correction factors to address two issues: incompleteness due to missing redshifts above the ECO luminosity limit, and incompleteness due to peculiar motions that extend outside the ECO redshift boundaries (affecting only Coma and two other large groups). We do not apply these completeness correction factors in our primary analysis, but we address how they alter results involving group-integrated H I mass in Section 5.

As detailed in Appendix, ECO DR3 includes an updated FoF group catalog used for comparison to the new G3 group finder (Section 3). The updated FoF group catalog reflects updates to ECO galaxy redshifts, ECO survey membership, and minor improvements to the FoF and false-pair splitting algorithms.

[15] These volume estimates update values from Eckert et al. (2016) based on minor corrections (∼1%) for cosmology using $\Omega_{\rm m,0} = 0.3$ and $\Omega_{\Lambda,0} = 0.7$.

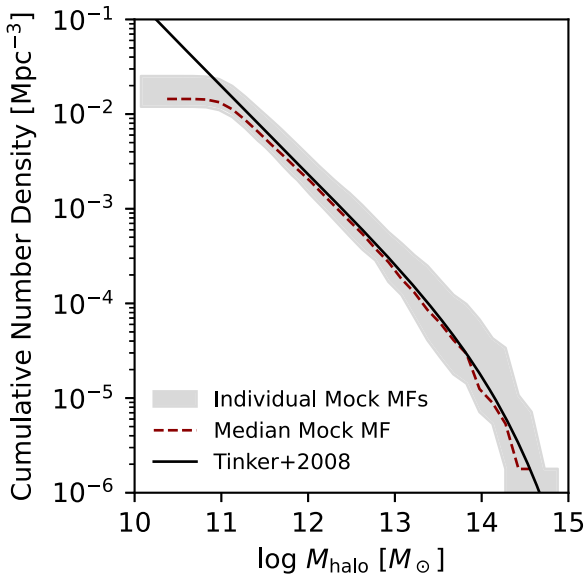


**Figure 1.** The median cumulative halo mass function of our eight fiducial mocks (red line) compared with the theoretical cumulative halo mass function from Tinker et al. (2008). The gray shading represents the range of values based on mock-to-mock variations reflecting cosmic variance.

For clarity, when comparing G3 and FoF groups throughout this paper, we use the "FoF" acronym to refer to our updated FoF algorithm (see Section A.5), while comparisons to the published group catalog of Eckert et al. (2017) are referenced with "E17 FoF."

### 2.5. Mock Catalogs

This work additionally requires synthetic mock catalogs to calibrate the group-finding procedure and test its correspondence to the underlying dark matter halos. Our analysis uses 32 custom mock catalogs that mirror the volume, shape, redshift limits, and luminosity function of the real ECO survey. The mocks were extracted from an *N*-body simulation assuming a Planck 2015 cosmology (Planck Collaboration 2016), with box size 180 $h^{-1}$ Mpc, $1080^3$ particles of individual mass $3.8\,h^{-1}\times10^8\,M_\odot$, and softening length $6h^{-1}$ kpc. The simulation begins at $z=99$ and is evolved to $z=0$ using GADGET-2 (Springel 2005). Halos were identified using ROCKSTAR (Behroozi et al. 2012), and halo masses were computed using spherical overdensities meeting a cosmology and redshift dependent virial threshold of approximately 337 times the mean background density. Figure 1 compares the cumulative halo mass function (CHMF) from each of our eight "fiducial" mocks (detailed below) with the theoretical CHMF from Tinker et al. (2008), which we use in Section 3 to estimate RESOLVE and ECO group halo masses via abundance matching. This comparison shows incompleteness in the mock mass functions below $\sim10^{11}\,M_\odot$, which affects our interpretations of group halo mass errors, described further in Section 4.

Synthetic galaxies were populated into the mock halos using the Zheng et al. (2007) halo occupation distribution method, and luminosities were assigned to synthetic galaxies to match the luminosity function of the real ECO survey and to match the conditional luminosity function of the Sloan Digital Sky Survey (York et al. 2000). Figure 2 illustrates distributions of real and mock galaxies' projected radii and velocities from group centers, with both real and mock groups determined using a simple FoF algorithm (Section 2.4). The distributions show a high degree of similarity, suggesting that mock galaxies are realistically populated in phase space.

Recent work has challenged two standard assumptions regarding the correspondence between the observed galaxy distribution and the underlying dark matter distribution in halos. First, Skibba et al. (2011) challenge the common assumption that a halo's most luminous galaxy, defined as the "central" in this work, sits at rest at the center of the halo. Second, both Guo et al. (2015) and Chen et al. (2018) challenge the common assumption that the halo velocity dispersion bias $b_v$, the ratio between the satellite galaxy velocity dispersion and the dark matter halo virial velocity dispersion (see definition in Tinker 2007), is equal to unity, when in reality, $b_v$ may depend on redshift or halo mass. To ensure that our group finder is robust to these biases, we test it in four distinct sets of eight mock catalogs. In a set of eight *fiducial* mocks, central galaxies are populated to the centers of halos and satellites are populated to random dark matter particle positions. In a set of eight *central-offset* mocks, central galaxies are instead populated to a random position within the halo virial radius, rather than the center of the halo gravitational potential well. Finally, in two sets of eight halo bias mocks, the velocity dispersion bias is set to $b_v=0.8$ or $b_v=1.2$, as opposed to the standard $b_v=1.0$. The eight individual mocks in a set have differing galaxy number densities, reflecting cosmic variance. The ECO mock catalogs are further described by M. Asad et al. (2023, in preparation).

## 3. The New G3 Group Finder

In high-redshift ($z\lesssim1$) surveys, data quality issues such as nonuniformity, shallowness, and incompleteness in the dwarf regime can lead to groups that are defined inconsistently in comparison to $z\sim0$ when employing conventional techniques such as FoF and halo abundance matching (HAM), which would hinder evolutionary analyses of group properties (e.g., of the group H I–halo mass relation). To consistently identify groups at $z\sim0$ and higher $z$ despite these data quality issues, we must avoid using halo mass information for group membership refinement, and we must avoid using dwarf galaxies as a primary basis for defining groups. Given the relevance of giant and dwarf regimes, a natural and physically meaningful mass scale that can inform group finding at $z\sim0$ is the gas-richness threshold scale of Kannappan et al. (2013). As shown in Figure 3, we use this scale at stellar mass $\sim10^{9.7}\,M_\odot$ to divide the galaxy population between giants and dwarfs at $M_r\sim-19.5$ mag at $z\sim0$; this divider might need to be modified at higher redshift.

The following subsections (Sections 3.1–3.5) describe the new RESOLVE Gas in Galaxy Groups (G3) group finder, its $z\sim0$ application to the highly complete and volume-limited RESOLVE and ECO surveys, and in particular, how its methodology can facilitate consistent group finding at $z\sim0$ and higher $z$ despite the aforementioned data quality issues. As conceptualized in the flow chart of Figure 4, our new group finder consists of four steps: (1) giant-only FoF to form initial giant-only groups, (2) iterative combination to merge and refine giant-only groups, (3) association of dwarfs into giant-only groups, and (4) iterative combination to construct dwarf-only groups among remaining dwarfs. This strategy allows us to use giant galaxies to form the initial basis of our groups, thereby mitigating the impact of bias from an incomplete selection of dwarf galaxies, but still allows us to identify dwarf-only groups

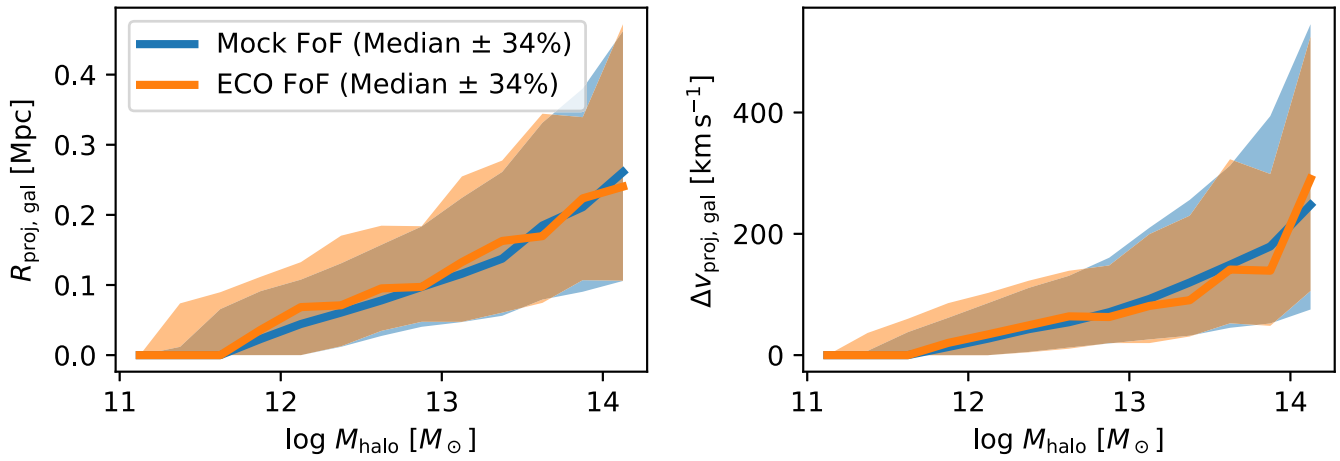


**Figure 2.** Distributions of real and mock FoF group members' projected relative radii ($R_{\rm proj,gal}$, left) and projected relative velocities ($\Delta v_{\rm proj,gal}$, right) from their average group centers as a function of group halo mass. Solid lines denote the median values, and shading encloses the 16th to 84th percentiles of each distribution. Orange represents real FoF groups from ECO (see Section 2.4), and blue represents groups extracted from the mocks using the same FoF algorithm. The mock data are drawn by combining the eight fiducial mocks into a single data set to minimize fluctuations due to cosmic variance. The agreement between real and mock FoF groups suggests that mock galaxies are populated realistically into halos.

in a final step (which may prove infeasible at higher $z$). Moreover, our new procedure avoids requiring group halo mass or radius information for membership refinement, instead relying on group $N_{\rm giants}$ or group-integrated luminosity, which are still available for incomplete, shallow, or nonuniform surveys. While Section 3.6 goes on to perform HAM to assign halo masses, this step is optional and does not affect the identified groups. The group finder as designed may be used for an incomplete and non-volume-limited survey for which HAM is infeasible.

### 3.1. Giant-only FoF

In the first step, we search for group "cores" by applying friends-of-friends (FoF) to only giant galaxies. Our giant-only FoF group finder follows the general procedure of Berlind et al. (2006). We select giant galaxies as having $M_r \leqslant -19.5$ and residing within $2530 \leqslant cz\,[\mathrm{km\ s^{-1}}] \leqslant 7470$ (ECO) or $4250 \leqslant cz\,[\mathrm{km\ s^{-1}}] \leqslant 7250$ (RESOLVE-B), as in Section 2.4. Galaxy $i$ and galaxy $j$ are linked if the projected distance between them satisfies

$$\frac{[d_M(z_i) + d_M(z_j)]}{2}(\alpha_{ij}) < b_\perp s \tag{1}$$

and if the line-of-sight distance satisfies

$$|d_C(z_j) - d_C(z_i)| < b_{||} s. \tag{2}$$

Here, $\alpha_{ij}$ is the angular separation between galaxies $i$ and $j$, $d_M$ is the comoving transverse distance in Mpc/radian at redshift $z$, $d_C$ is the line-of-sight comoving distance at redshift $z$, $b_{||}$ and $b_\perp$ are the line-of-sight and transverse linking constants, and $s$ is the mean separation of giant galaxies. We adopt linking constants of $b_{||} = 1.1$ and $b_\perp = 0.07$ following the recommendation of Duarte & Mamon (2014) for the study of galaxy environment. This choice is consistent with that of Eckert et al. (2016) in making the RESOLVE and ECO FoF catalogs. For the ECO survey, we estimate $s \sim 4.84$ Mpc from the group-finding volume, which we also use for RESOLVE-B to minimize cosmic variance issues (see Section 2.4). Other $z \sim 0$ surveys can use the ECO value of $s$, as ECO has overall number density similar to the entire SDSS (Eckert et al. 2016). Beyond $z \sim 0$, however, it will be necessary to remeasure $s$.

### 3.2. Giant-only Group Merging

A potential problem with giant-only FoF is that it may tend to produce multiple giant-only groups that are subgroups within one true system and that would have been linked if dwarfs had been included. To address this issue, we calibrate the typical transverse and line-of-sight offsets of group members versus $N_{\rm giants}$, the number of giant galaxies in a group, for use in both merging giant-only groups that really represent one true group and associating dwarfs to these groups in step 3 (Section 3.3 below). We choose to calibrate using $N_{\rm giants}$ because this quantity is easily measured even for surveys for which halo information is not easily derived.

The calibrations are derived as follows. We first determine the group centers of giant-only groups by averaging the positions of giant galaxies in Cartesian coordinates centered on the observer, i.e.,

$$\mathcal{O}_{\rm avg} = \frac{1}{N_{\rm giants}} \sum_{i}^{N_{\rm giants}} \begin{bmatrix} d_c(z_i)\sin\theta_i \cos\phi_i \\ d_c(z_i)\sin\theta_i \sin\phi_i \\ d_c(z_i)\cos\theta_i \end{bmatrix}, \tag{3}$$

where $d_c(z_i)$ is the line-of-sight comoving distance, $\phi_i = \mathrm{RA}_i$, $\theta_i = \pi/2 - \mathrm{DEC}_i$, and $N_{\rm giants}$ is the number of giant galaxies in each group. Section 4.3 describes this choice of group center definition in more detail. Using this group center, we are able to calculate the projected radius $R_{\rm proj,gal}$ and relative velocity $\Delta v_{\rm proj,gal}$ of each giant galaxy relative to its group $\mathcal{O}_{\rm avg}$. In each $N_{\rm giants}$ bin, we compute the median projected radius $R_{\rm proj}$ and median relative velocity $\Delta v_{\rm proj}$ for all giants in all FoF groups in that bin. Using bootstrapped errors as weights, we fit these medians as a function of $N_{\rm giants}$ to the simple models

$$R_{\rm proj}^{\rm fit} = a_R \log_{10}(b_R N_{\rm giants} + 1) \tag{4}$$

and

$$\Delta v_{\rm proj}^{\rm fit} = a_v \log_{10}(b_v N_{\rm giants} + 1). \tag{5}$$

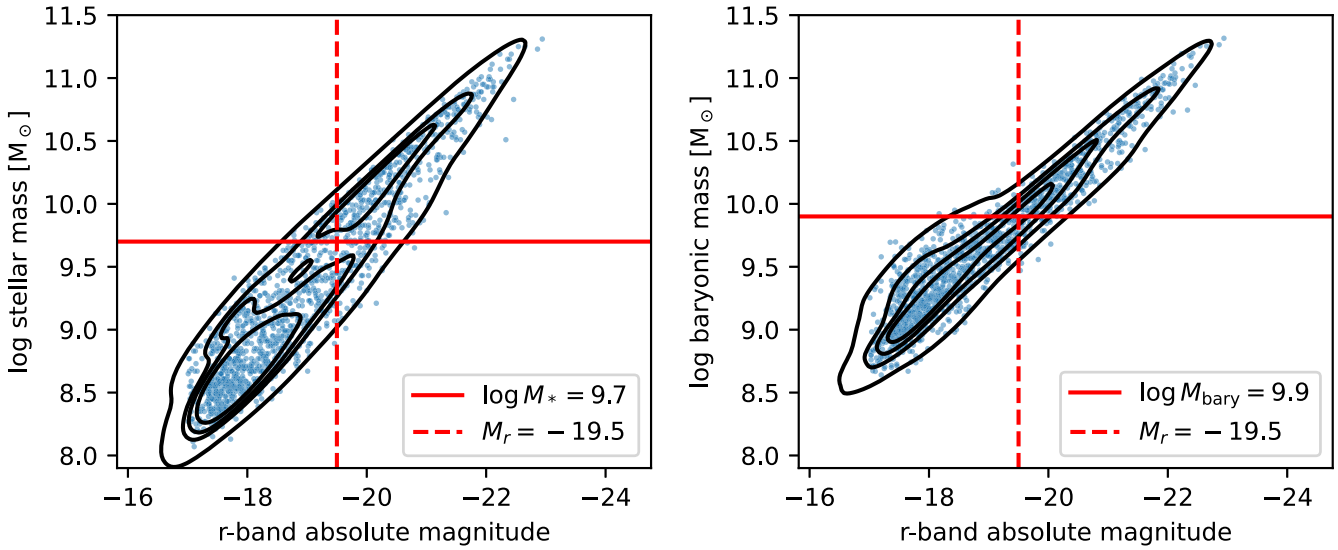


**Figure 3.** Scatter plots of stellar mass (left) and cold baryonic mass (right, $M_* + 1.4M_{\rm H\,I}$) as functions of absolute $r$-band magnitude for RESOLVE galaxies. The red lines illustrate our definition of the dwarf–giant divide in luminosity, stellar mass, and cold baryonic mass. We use this scale to divide our group finder into four discrete steps, as discussed in Section 3.

The best-fit parameters to Equations (4) and (5) are given in Table 1. Figure 5 shows $R_{\rm proj}$ and $\Delta v_{\rm proj}$ versus $N_{\rm giants}$ and the corresponding fits, up to $N_{\rm giants} = 20$.

With these relations in hand, we iteratively combine neighboring pairs of initial giant-only groups that likely represent one real system. This "iterative combination" process is performed as follows:

1. Compute $N_{\rm giants}$ and $\mathcal{O}_{\rm avg}$ for each giant-only group. In the first pass, these correspond to the initial giant-only groups from FoF.
2. Use a k-d tree (Bentley 1975; Virtanen et al. 2020) to identify nearest-neighbor giant-only groups.
3. For every nearest-neighbor pair of giant-only groups $i$ and $j$, evaluate whether the pair should be merged into a single, larger giant-only group:
   (a) Compute the total number of giants for the pair: $N^{\rm pair}_{\rm giants} = N_{{\rm giants},i} + N_{{\rm giants},j}$.
   (b) Evaluate the boundaries $3R^{\rm fit}_{\rm proj}(N^{\rm pair}_{\rm giants})$ and $4\Delta v^{\rm fit}_{\rm proj}(N^{\rm pair}_{\rm giants}) + 200\ {\rm km\ s^{-1}}$.
   (c) If the projected distance and relative velocity between the group centers of $i$ and $j$ are smaller than these boundaries, merge $i$ and $j$ into a single, larger giant-only group.
4. Repeat from step 1 until the giant-only group catalog has converged, i.e., when giant-only groups are no longer changing between iterations.

The choices of the boundaries at $3R^{\rm fit}_{\rm proj}$ and $4\Delta v^{\rm fit}_{\rm proj} + 200\ {\rm km\ s^{-1}}$ are derived with mocks in Section 4.2. The result of this procedure is a catalog of giant-only groups. In the next step, we associate dwarf galaxies into these giant-only groups using the same $R^{\rm fit}_{\rm proj}$ and $\Delta v^{\rm fit}_{\rm proj}$ relations.

### 3.3. Associating Dwarfs into Giant-only Groups

Our next task is to refine these giant-only groups to include dwarf galaxies that were initially excluded. The design of this process is similar to that of several published techniques that perform group membership refinement. For example, Yang et al. (2005) iteratively refine seed groups based on inferred halo masses and radii, and Hess & Wilcots (2013) use an H I-proximity matching technique to include H I-detected dwarfs in optically selected FoF group catalogs. To associate dwarf galaxies to the giant-only groups found in Section 3.1, we apply the relations illustrated in Figure 5. We assign a dwarf galaxy to a giant-only group if it falls within association distances $3R^{\rm fit}_{\rm proj}$ and $4\Delta v^{\rm fit}_{\rm proj} + 200\ {\rm km\ s^{-1}}$ of $\mathcal{O}_{\rm avg}$, as used for combining giant-only groups in the previous section. These relations can be extrapolated to $N_{\rm giants} = 1$, allowing us to compute association distances around $N_{\rm giants} = 1$ giant-only groups, which otherwise have no obvious group radius or velocity boundaries.

If a given dwarf galaxy satisfies this condition for multiple giant-only groups, we assign it to the group for which the quantity

$$\sqrt{\left(\frac{R_{\rm proj,gal}}{3R^{\rm fit}_{\rm proj}}\right)^2 + \left(\frac{\Delta v_{\rm proj,gal}}{4\Delta v^{\rm fit}_{\rm proj} + 200\ {\rm km\ s^{-1}}}\right)^2} \tag{6}$$

is minimized. Other $z \sim 0$ surveys can use these parameters and our fits to Equations (4) and (5) to determine association distances, but beyond $z \sim 0$ it will be necessary to refit the relations and reoptimize the parameters. The groups resulting from this technique may contain both giants and dwarfs or just giants, so we term them "giant-hosting" groups.

### 3.4. Identifying Dwarf-only Groups

By associating dwarfs to giant-only groups, we have recovered most groups, but have overlooked any dwarf-only groups that do not contain a giant. In our last group-finding step, we look to identify any remaining dwarf-only systems. This step is only robust for a complete and volume-limited survey such as RESOLVE or ECO.

To identify dwarf-only groups, we iteratively combine dwarfs using on-sky and line-of-sight association distances determined by extrapolation from association distances for giant-hosting groups. However, as our previous relations

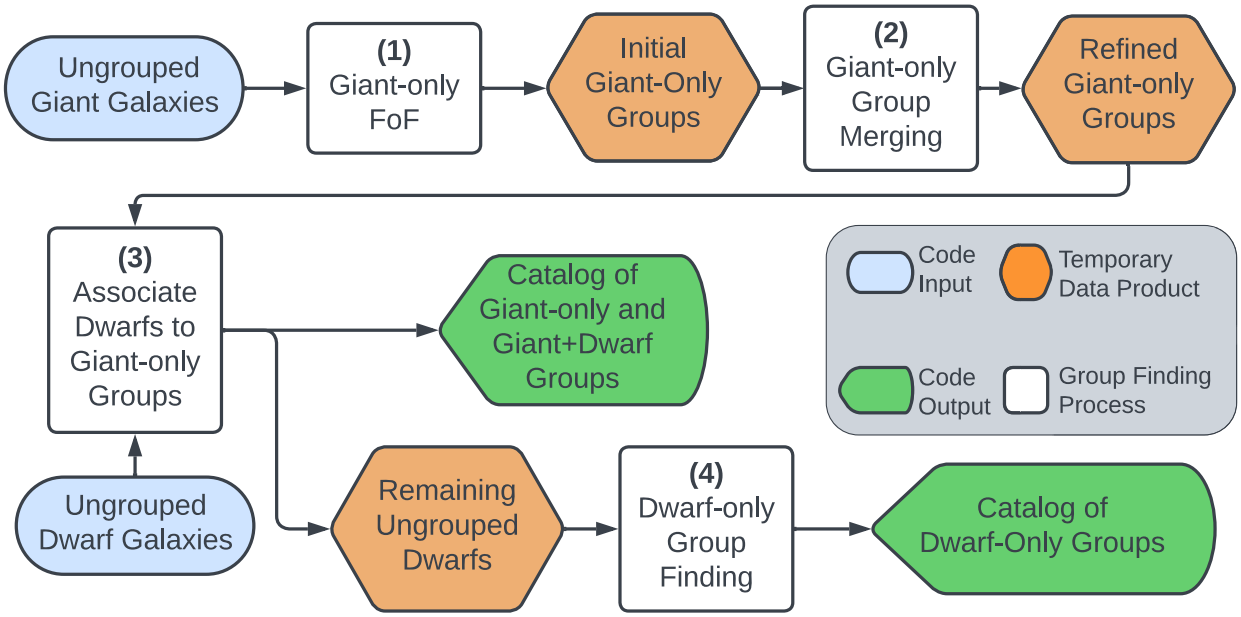


**Figure 4.** Flow chart conceptualizing the new RESOLVE Gas in Galaxy Groups (G3) group-finding algorithm, consisting of four distinct group-finding processes: (1) giant-only FoF to construct initial giant-only groups, (2) giant-only merging to refine giant-only groups, (3) dwarf association to construct groups containing both giants and dwarfs, and (4) identification of dwarf-only groups.

**Table 1**
Best-fit Values to Equations (4), (5), (7), and (8)

| Parameter | Unit | Value | Error |
|---|---|---|---|
| $a_R$ | Mpc | $3.06 \times 10^{-1}$ | $0.49 \times 10^{-1}$ |
| $b_R$ | | $4.16 \times 10^{-1}$ | $1.1 \times 10^{-1}$ |
| $a_v$ | km s$^{-1}$ | $3.45 \times 10^{2}$ | $0.42 \times 10^{2}$ |
| $b_v$ | | $1.70 \times 10^{-1}$ | $0.28 \times 10^{-1}$ |
| $\alpha_R$ | Mpc | $3.42 \times 10^{-2}$ | $0.14 \times 10^{-2}$ |
| $\beta_R$ | mag$^{-1}$ | $5.10 \times 10^{-1}$ | $0.14 \times 10^{-1}$ |
| $\alpha_v$ | km s$^{-1}$ | $1.97 \times 10^{1}$ | $0.10 \times 10^{1}$ |
| $\beta_v$ | mag$^{-1}$ | $4.16 \times 10^{-1}$ | $0.19 \times 10^{-1}$ |

**Notes.** The top four rows correspond to parameters of Equations (4) and (5), while the bottom four rows correspond to Equations (7) and (8).

defined association distances as functions of $N_{\rm giants}$, which is by definition zero for dwarf-only groups, we calibrate new relations as a function of group-integrated luminosity $M_{r,\rm tot}$. Group-integrated luminosity is an observable proxy for group halo mass, and giant-hosting and dwarf-only groups show similar $R_{\rm proj,gal}$ distributions as a function of halo mass in mock catalogs (see Figure 6).

Figure 7 shows the results of the new calibration. We fit the median projected radii and velocities against the integrated luminosity of giant-hosting groups using the models

$$R^{\rm gi,dw}_{\rm proj,fit} = \alpha_R \exp\left(-\beta_R[M_{r,\rm tot} + 19.5]\right) \quad (7)$$

and

$$\Delta v^{\rm gi,dw}_{\rm proj,fit} = \alpha_v \exp\left(-\beta_v[M_{r,\rm tot} + 19.5]\right). \quad (8)$$

The best-fit parameters for Equations (7) and (8) are given in Table 1.

Dwarf-only group finding is performed by iterative combination as described in Section 3.2. In the first pass, we consider each ungrouped dwarf an $N_{\rm dwarfs} = 1$ seed group. Each iteration then combines dwarf-only groups, given boundaries $2R^{\rm gi,dw}_{\rm proj,fit}$ and $4\Delta v^{\rm gi,dw}_{\rm proj,fit} + 100$ km s$^{-1}$, until the dwarf-only groups converge. The scaling parameters of 2, 4, and 100 km s$^{-1}$ are shown to be optimal using mock catalogs in Section 4.2.

The result of the iterative combination process is a catalog of dwarf-only groups. As shown in the shaded and hatched regions of Figure 8, the G3 algorithm identifies fewer $N_{\rm dwarfs} > 1$ dwarf-only groups than were found in the E17 FoF group catalog. To compare directly with the dwarf-only groups identified by Stierwalt et al. (2017), we computed the observational projected radii of our dwarf-only groups as the largest projected separation between any two group members. The typical projected radius in Stierwalt et al. (2017) is ∼30 kpc, whereas we find median values of 68 kpc, 108 kpc, and 45 kpc for the G3 groups, E17 FoF groups, and mock true dwarf-only groups, respectively. The G3 dwarf-only groups represent the full population of dwarf-only groups, whereas the seven dwarf-only groups of Stierwalt et al. (2017) were selected specifically to be compact. Conversely, the dwarf associations of Tully et al. (2006) are generally larger than the G3 dwarf-only groups, as these associations contain 4–6 dwarfs on scales of hundreds of kiloparsecs and may consist of multiple dark matter halos (see Kourkchi & Tully 2017). Some of the E17 FoF dwarf-only groups may resemble such associations, as E17 FoF yields a higher median projected radius of 108 kpc and finds more dwarf triplets and quintets than does the G3 algorithm.

### 3.5. Comparison to E17 Group Catalog

Figure 8 compares the multiplicity function of groups identified by the new G3 algorithm detailed above to the multiplicity function corresponding to the published E17 group catalog, which was built from FoF plus a special procedure for breaking false pairs. The E17 false-pair splitting algorithm reduced the number of pairs in ECO from ∼800 to 568. The G3 algorithm intrinsically identifies 666 galaxy pairs, including pairs with giants, in ECO. Model fits to the G3 and E17 FoF multiplicity functions show that G3 follows a decaying exponential $[3.9 \pm 0.2]\exp([-0.13 \pm 0.01]N_{\rm galaxies})$, while E17 FoF

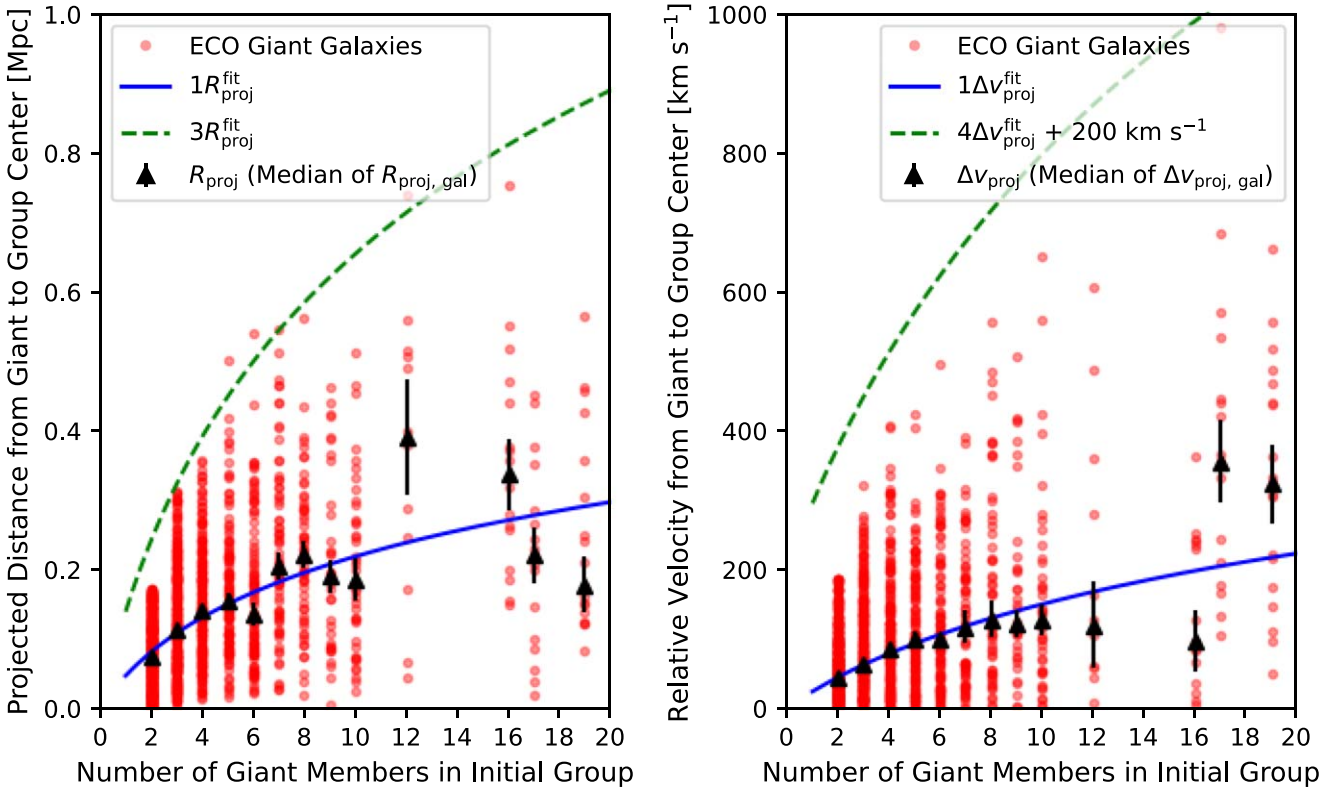


**Figure 5.** Plots illustrating $R_{\rm proj}$ (left) and $\Delta v_{\rm proj}$ (right) for giant-only FoF group galaxies as a function of $N_{\rm giants}$. The triangles represent the median relative projected radii and velocities of giant galaxies in $N_{\rm giants} > 1$ FoF groups (red points), as a function of $N_{\rm giants}$. The blue lines show our best fit to the medians using Equations (4) and (5), and the green dashed–dotted lines show the boundaries we use for merging giant-only FoF groups and associating dwarfs to these groups (Section 3.3). For visual clarity, $N_{\rm giants} \geqslant 20$ groups are omitted from the plot, but the median values for these groups are consistent with the trend shown.

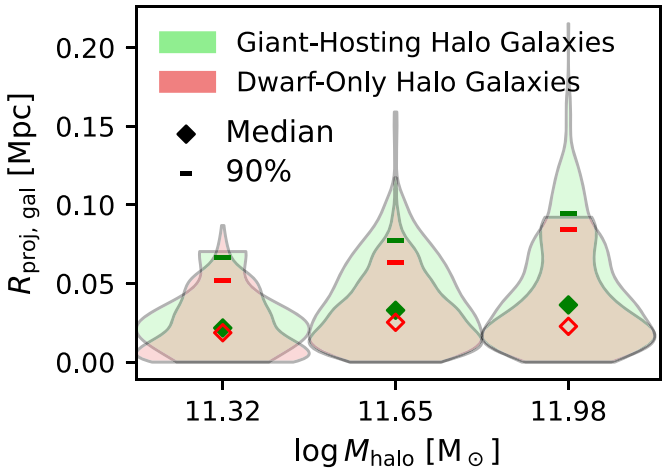


**Figure 6.** Violin plot illustrating distributions of $R_{\rm proj,gal}$ conditioned on group halo mass for true giant-hosting groups and true dwarf-only groups from the mock catalogs. Diamonds represent median quantities, and solid bars represent 90th percentiles. Violin widths represent the frequency distribution of $y$-axis data points in each horizontal bin.

follows $[4.3 \pm 0.2]\exp([-0.19 \pm 0.01]N_{\rm galaxies})$. These best-fitting values suggest similar overall normalization but a faster ramp-down for E17 FoF. Compared to G3, the E17 FoF multiplicity function shows a deficit of groups with more than ∼10 galaxies. This result is consistent with our finding that high-mass FoF groups are less complete on average than high-mass G3 groups (see Section 4 below). However, when we compare the G3 and FoF group $N_{\rm galaxies}$ distributions in fixed halo mass bins, two-sample Kolmogorov–Smirnov tests yield only weak or insignificant probability that the G3 and FoF $N_{\rm galaxies}$ distributions are distinct at fixed halo mass. We thus conclude that the G3 halo occupation distribution is not significantly different from the FoF halo occupation distribution.

### 3.6. Halo Mass Estimation

Although our new group finder has been designed to avoid requiring halo abundance matching (e.g., for group membership refinement), we choose to apply HAM to the RESOLVE and ECO groups to obtain additional reference data, because we know the surveys to be complete and volume-limited. Following Blanton & Berlind (2007), we build a one-to-one monotonic relationship between group luminosity and halo mass (Figure 9) by matching the cumulative number density distribution of groups above a given luminosity to the theoretical cumulative number density distribution of dark matter halos above a given mass. We use the theoretical halo mass function of Tinker et al. (2008) and adopt a mean background overdensity of $\Delta_{\rm vir} = 337$ to compute group halo virial masses and radii, matching the convention in our mocks. For RESOLVE-B groups, we perform abundance matching with a RESOLVE-B analog version of ECO, as described in Section 2.4 and also by Eckert et al. (2016). This matching provides a second group luminosity–halo mass relation reaching down to $M_r = -17.0$, which we interpolate to obtain halo masses for RESOLVE-B groups. We note that the ECO $M_{\rm halo}$–$M_{r,\rm tot}$ relation can be interpolated to estimate halo masses for other $z \sim 0$ surveys for which HAM is not viable, because of ECO's similar density to the overall $z \sim 0$ mean density in the full SDSS (Eckert et al. 2016). The ECO DR3 group

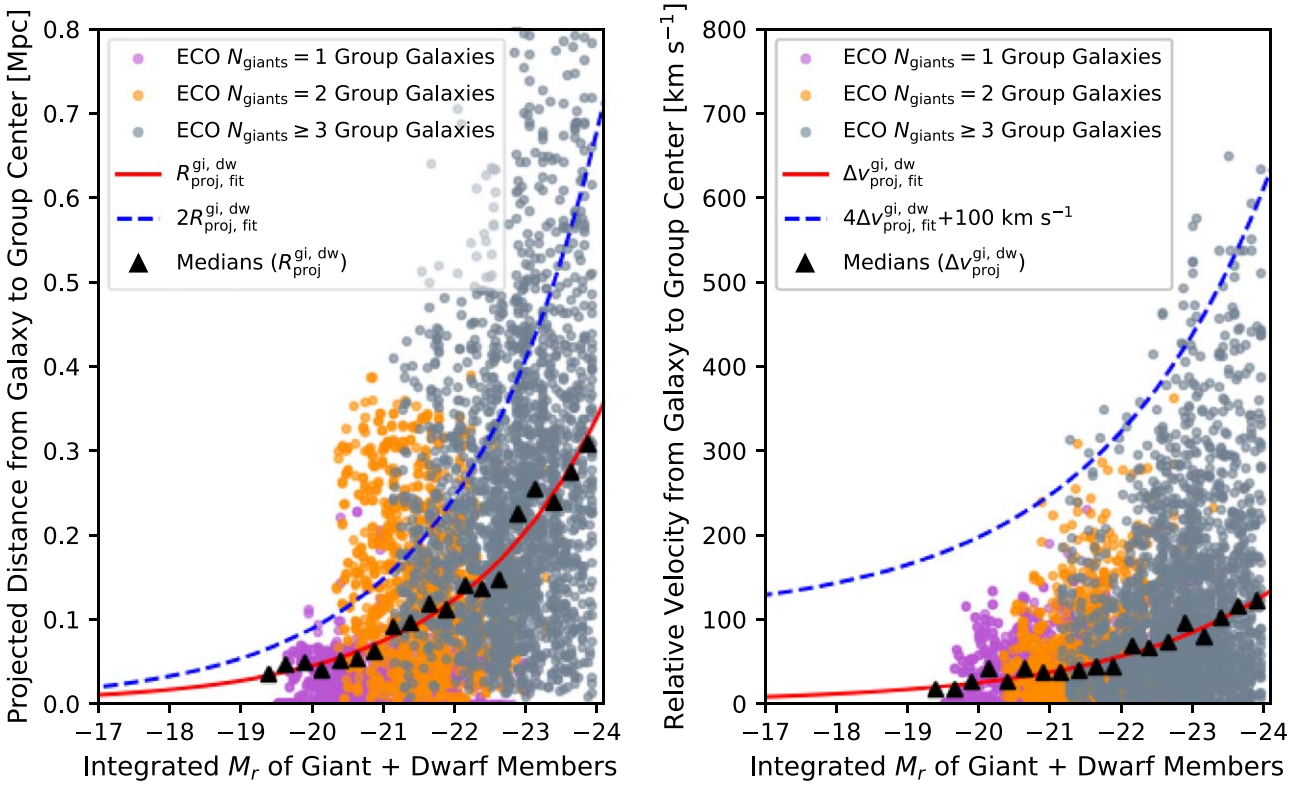


**Figure 7.** Calibration of boundaries used to perform iterative combination in dwarf-only group finding. Data points represent the on-sky distances (left) and relative velocities (right) of giant or dwarf galaxies in giant-hosting groups (i.e., not dwarf-only groups) and are colored by $N_{giants}$. Black triangles represent the medians of these radii and velocities in different group luminosity bins. The red lines show fits to the black triangles as a function of group-integrated luminosity. The blue dashed lines are the scaled fits, which define the group boundaries for dwarf-only group finding, as described in Section 3.4.

catalog in Section A.5 tabulates group-integrated luminosity and group-integrated giant galaxy luminosity for this purpose.

One possible alternative to HAM is dynamical mass estimation, and previous work by E17 has shown that dynamical mass estimation is superior to HAM in the high-mass regime of group $N_{galaxies} > 7$. However, our surveys are dominated by small groups (e.g., 98% of ECO groups have $N_{galaxies} \leqslant 7$). For simplicity, we use a single set of group-finding parameters that do not vary with group $N_{galaxies}$ (see Section 4.2). With this particular set of group-finding parameters, our mock catalog analysis suggests that our dynamical mass estimates are not notably more accurate than HAM. Thus, we choose to use HAM halo masses for the remainder of this paper.

## 4. Optimizing the G3 Group Finder

The new G3 group-finding algorithm presented in Section 3 employs methods for merging giant-only groups, associating dwarfs into giant-only groups, and constructing dwarf-only groups. These methods have tunable parameters, namely the scalar multipliers for each of $R^{fit}_{proj}$, $\Delta v^{fit}_{proj}$, $R^{gi,dw}_{proj,fit}$, and $\Delta v^{gi,dw}_{proj,fit}$, as well as the vertical offsets to $\Delta v^{fit}_{proj}$ and $\Delta v^{gi,dw}_{proj,fit}$. For future G3 science, it is crucial that these parameters be selected to ensure accurate recovery of groups. However, there is no single set of parameters for the G3 algorithm that can perfectly recover groups, and previous work has shown that optimal group finding may depend on one's scientific goal (e.g., Berlind et al. 2006; Duarte & Mamon 2014). Consequently, the users of a group-finding algorithm must review the range of parameters that can be passed to an algorithm, and select the "best" values based on relative trade-offs. In this section, we optimize the G3 group finder by testing many sets of possible parameters in mock catalogs, then visualizing and evaluating their relative performance in terms of purity, completeness, and halo mass error. With the group finder optimized, we then apply it to mocks to confirm that overall purity, completeness, and halo mass error distributions are robust to issues of halo velocity bias, central galaxy location, and cosmic variance.

### 4.1. Metrics of Group-finding Performance

To quantify group-finding performance, we use our mock catalogs to compute metrics of purity, completeness, and halo mass error by comparing "true" groups to groups recovered by the G3 algorithm. In our mock catalogs, "true" groups are defined by occupancy of a common dark matter halo.

For this analysis, we compute purity and completeness statistics because they directly evaluate how well groups are recovered in terms of fragmentation and overmerging (e.g., see Yang et al. 2005; Gerke et al. 2007). To map each identified group $g$ to a primary true group (halo) $h$ in the mock catalog, we define $h$ as the mode of halo ID numbers of galaxies classified as part of $g$. In cases of multiple modes, we use the mode that contains the most luminous member in the $r$ band. The purity and completeness of group $g$ is then calculated as $P = N_s/N_g$ and $C = N_s/N_h$ with the following definitions: $N_g$ is the number of galaxies in $g$, $N_h$ is the number of galaxies in $h$, and $N_s$ is the number of selected galaxies, i.e., galaxies in $g$ that are also in $h$.

We note that this mapping of groups to halos ($g \rightarrow h$) is not symmetric or one-to-one, and that due to group-finding errors,

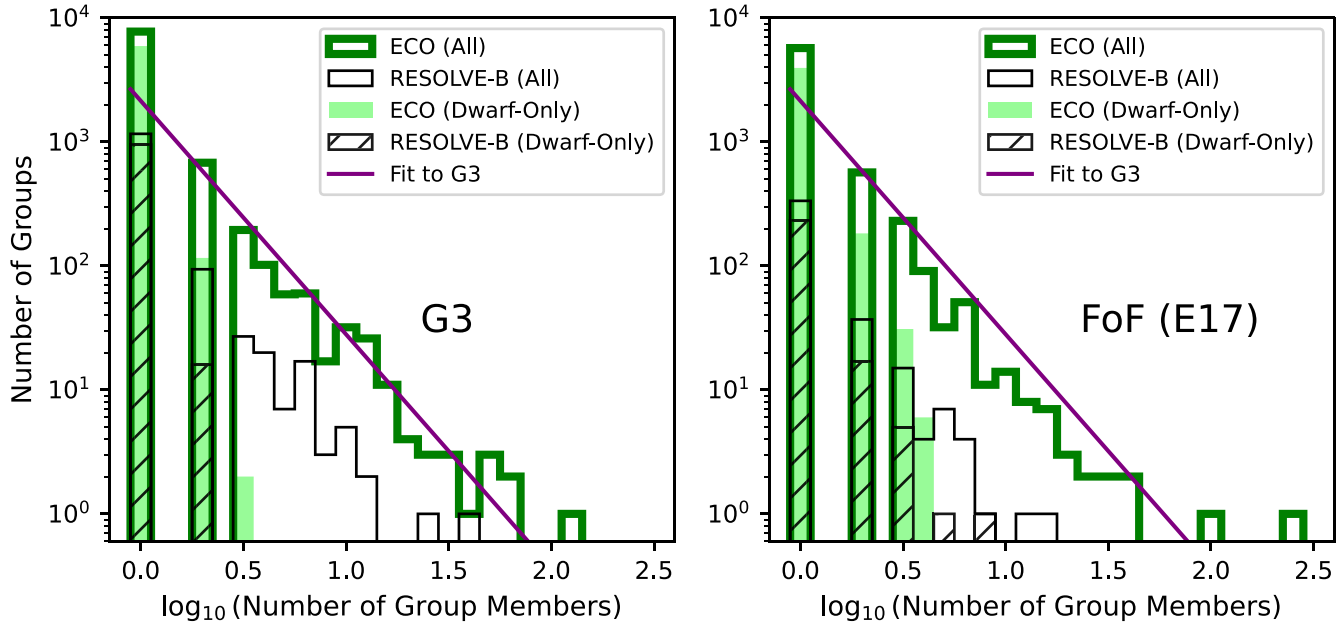


**Figure 8.** Multiplicity functions for groups in RESOLVE-B and ECO. The left panel shows groups constructed with the G3 algorithm described in Sections 3.1–3.4. The right panel shows groups as published in Eckert et al. (2017), constructed with friends-of-friends and false-pair splitting. In both panels, the shaded (ECO) and hatched (RESOLVE-B) regions highlight dwarf-only groups. The purple line is a model fit to the ECO G3 multiplicity function and is shown for reference in both panels.

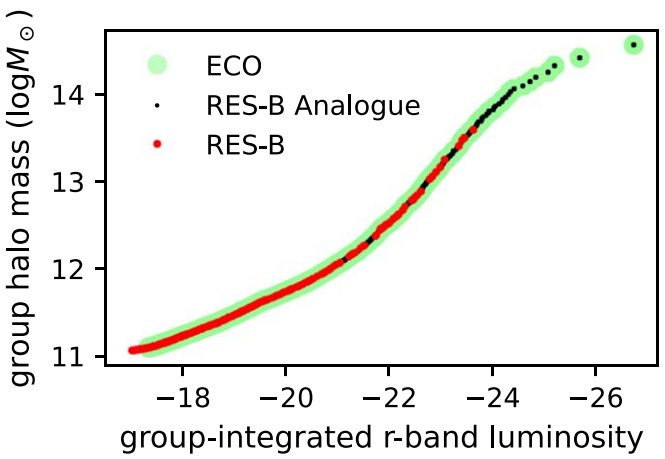


**Figure 9.** Halo masses in the RESOLVE and ECO G3 group catalogs as a function of group-integrated $r$-band luminosity. The masses are determined using halo abundance matching as described in Section 3.6.

the reversed mapping ($h \rightarrow g$) is likely to produce a slightly different set of objects. To address this issue, we compute purity and completeness as defined above for both mappings, using $P_g$ and $C_g$ to denote the $g \rightarrow h$ mapping, and likewise $P_h$ and $C_h$ for the $h \rightarrow g$ mapping. In the $h \rightarrow g$ mapping, $N_s$ is the number of galaxies in $h$ that are also in $g$. By definition, $N_{\rm galaxies} = 1$ groups always have $P_g = 1$, but their $P_h$ and completeness values will reflect any fragmentation errors.

Finally, because the main goals of the G3 project involve examining group scaling relations at fixed halo mass, we also assess halo mass error as part of our optimization. For each galaxy $i$ in a mock catalog, we define $E_i = |\log M_{{\rm HAM},i} - \log M_{{\rm true},i}|$ as the absolute halo mass deviation. Because $E_i$ is computed on a galaxy-by-galaxy basis, individual $E_i$ values directly reflect fragmentation or overmerging errors (e.g., a field galaxy incorrectly grouped into a cluster might have $\log M_{\rm HAM}[M_\odot] = 14$ but $\log M_{\rm true}[M_\odot] = 11$, yielding $E_i = 3$ dex). This definition also means that each group has a set of absolute halo mass deviations $\{E_i\}$ whose length is the number of group members. The halo mass error of the group can be computed as $\tilde{E} = {\rm median}(\{E_i\})$. For the purpose of optimizing the group finder, it is helpful to have a single statistic that describes the typical halo mass error in a mock group catalog. We thus define the typical halo mass error $\mu_{\rm HME}$ as the median of the set of median group $E_i$ values, i.e., $\mu_{\rm HME} = {\rm median}(\{\tilde{E}_1, \tilde{E}_2, ..., \tilde{E}_k\})$ for a mock group catalog that consists of $k$ identified groups. We use this quantity $\mu_{\rm HME}$ for our optimization in the following sections, but we note that we tested different definitions for $\mu_{\rm HME}$ (e.g., the weighted median of all $E_i$ values where the weight for each galaxy is proportional to its group $N_{\rm galaxies}$) and the optimal parameters did not change.

### 4.2. Optimizing Algorithm Parameters

In this subsection, we use the metrics defined above to determine the optimal parameters for our group finder. To do so, we apply the group finder to a single fiducial mock catalog across a six-dimensional grid of possible parameters, with each of the grid coordinates representing one of the six free parameters. For example, the coordinate (1, 2, 300, 4, 5, 600) would correspond to group finding with $1R_{\rm proj}^{\rm fit}$ and $2\Delta v_{\rm proj}^{\rm fit} + 300$ km s$^{-1}$ for giant-only group merging and dwarf association, plus $4R_{\rm proj,fit}^{\rm gi,dw}$ and $5\Delta v_{\rm proj,fit}^{\rm gi,dw} + 600$ km s$^{-1}$ for dwarf-only group finding. At each point on this grid, we construct a mock group catalog with the given parameters and use it to compute the mean purity statistics $\bar{P}_g$ and $\bar{P}_h$, mean completeness statistics $\bar{C}_g$ and $\bar{C}_h$, and median halo mass error $\mu_{\rm HME}$. Given that $N_{\rm galaxies} = 1$ groups dominate our mock catalogs, and because these $N_{\rm galaxies} = 1$ groups have high purity and completeness independent of group-finding parameters, the median purity and completeness statistics are insensitive to variations on the group-finding parameters. For

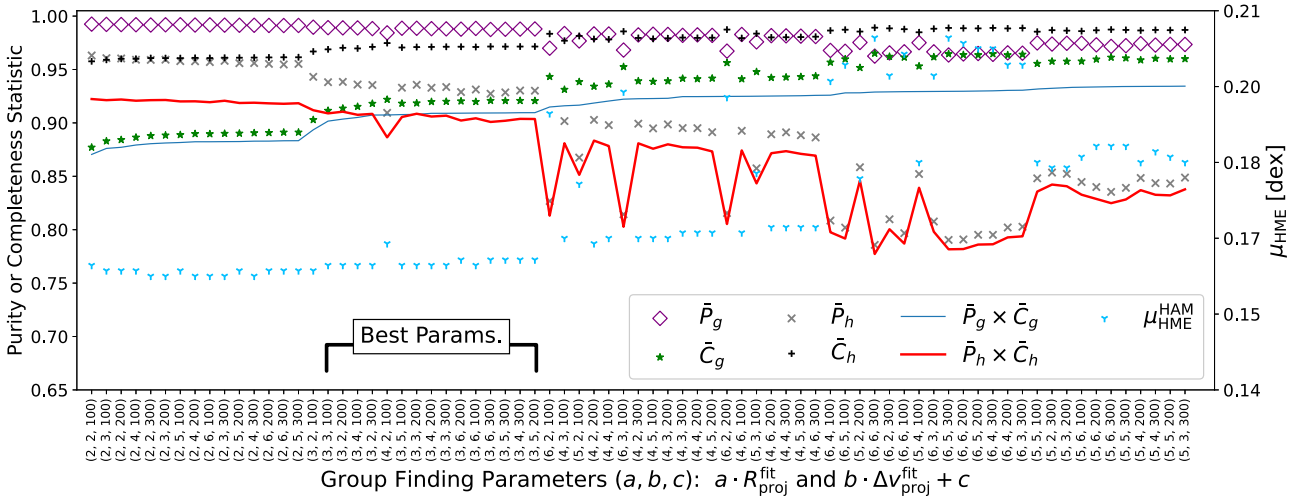


**Figure 10.** Scatter plot comparing group-finding performance for different parameter choices for the giant-only merging and dwarf association techniques presented in Sections 3.2–3.3. Each column corresponds to a different set of parameters for $R_{\rm proj}^{\rm fit}$ and $\Delta v_{\rm proj}^{\rm fit}$; e.g., (1,2,300) corresponds to $1R_{\rm proj}^{\rm fit}$ and $2\Delta v_{\rm proj}^{\rm fit} + 300$ km s$^{-1}$. The plotted markers in each column show the mean purity and completeness (left $y$-axis) as well as halo mass error (right $y$-axis) statistics corresponding to each set of candidate parameters. Solid red and blue lines show the product of purity and completeness for the $g \rightarrow h$ and $h \rightarrow g$ mappings, respectively. The black bracket marks the range of parameters that best balances purity and completeness and yields low halo mass error.

this reason, we use mean values for purity and completeness to better detect changes in group-finding performance when switching between candidate parameters. The result is a table of candidate group-finding parameters and their group-finding performance metrics. Ideally, we want to use the parameters that achieve purity and completeness of unity and $\mu_{\rm HME} = 0$, but in reality, these metrics trade-off with one another (e.g., purity for completeness) and compromises must be made when selecting the best parameters.

We start by focusing on the three free parameters that scale $R_{\rm proj}^{\rm fit}$ and $\Delta v_{\rm proj}^{\rm fit}$ for giant-only group merging and dwarf association. In Figure 10, we illustrate group-finding statistics for ∼75 different candidate sets of these three parameters, plotted in vertical columns. Each column is not unique, because the free parameters used in dwarf-only group finding are not yet specified. The markers in each column thus represent median values over the superset of optimization runs using all tested dwarf-only group-finding parameters. We find that variations due to different choices of dwarf-only group-finding parameters are typically ∼1%. The black bracket highlights the range of parameters in which the group finder achieves a compromise on increasing $\bar{C}_g$ and decreasing $\bar{P}_h$, while simultaneously reaching high $\bar{P}_g$ and $\bar{C}_h$ (>90%) and low halo mass error ($\mu_{\rm HME} \sim 0.16$ dex). Our preferred parameters are selected at the middle of this range and correspond to $3R_{\rm proj}^{\rm fit}$ and $4\Delta v_{\rm proj}^{\rm fit} + 200$ km s$^{-1}$, as introduced in Section 3.2. We note that, because small groups greatly outnumber large groups in volume-limited data sets, our group-finding parameters are not optimized for large groups (e.g., restricting to $N_{\rm galaxies} > 7$, as would be used for dynamical masses, yields different optimal parameters.)

Our next task is to select the other three free parameters that scale $R_{\rm proj,fit}^{\rm gi,dw}$ and $\Delta v_{\rm proj,fit}^{\rm gi,dw}$ for the iterative combination of dwarf-only groups. To do so, we examine a subset of the grid with fixed values corresponding to $3R_{\rm proj}^{\rm fit}$ and $4\Delta v_{\rm proj}^{\rm fit} + 200$ km s$^{-1}$. Figure 11 shows group-finding statistics for $N_{\rm dwarfs} > 1$ dwarf-only groups given ∼75 choices of dwarf-only group-finding parameters plotted on the horizontal axis. The rationale for excluding $N_{\rm dwarfs} = 1$ groups is that $\bar{P}_g$ and $\bar{P}_h$ are uniformly high when dwarf singles are considered, as the vast majority of dwarf singles are perfectly identified as $N_{\rm dwarfs} = 1$ and have $P_g = 1$ and $P_h = 1$, which causes the optimization to focus on maximizing completeness regardless of whether the small minority of $N_{\rm dwarfs} > 1$ groups are impure. We find that scaling parameters corresponding to $2R_{\rm proj,fit}^{\rm gi,dw}$ and $4\Delta v_{\rm proj,fit}^{\rm gi,dw} + 100$ km s$^{-1}$, as introduced in Section 3.4, jointly maximize all purity and completeness statistics and minimize $\mu_{\rm HME}$ for $N_{\rm dwarfs} > 1$ dwarf-only groups.

### 4.3. Testing Group Center Definitions

A crucial design choice of the G3 group finder is our use of a simple average group center (Equation (3)) for merging giant-only groups, associating dwarfs, and constructing dwarf-only groups. We arrived at this choice after considering two common alternatives. First, we ruled out a mass- or luminosity-weighted group center on the basis that such a definition would be difficult to apply consistently in flux-limited, nonuniform, or incomplete surveys.

Second, we more seriously considered a popular alternative choice, which is to assume the group center is located at the position of what we call the "central" (equivalent to the brightest cluster galaxy or BCG in other work). Early work by Zabludoff & Mulchaey (1998) showed that the average group center and central-based group center were indistinguishable for galaxy groups with halo masses above $10^{13.6}\,M_\odot$ (approximately $N_{\rm giants} = 5$). However, more recent work has shown that a substantial fraction of centrals, 40%–60%, are not located at group centers, for group halo masses down to $10^{12}\,M_\odot$ (van den Bosch et al. 2005; Skibba et al. 2011). The validity of a central-defined group center likely depends on halo virialization state and may depend on halo mass; thus, if we use a central-defined center, we may risk systematic group-finding errors (e.g., fragmentation).

Nonetheless, we tested three algorithms: (i) a simple average as in Equation (3), which is implemented as a giant-only average for giant-only group finding and dwarf association, (ii) a central-based group center, and (iii) an adaptive definition that smoothly varies from a simple average to a central-based

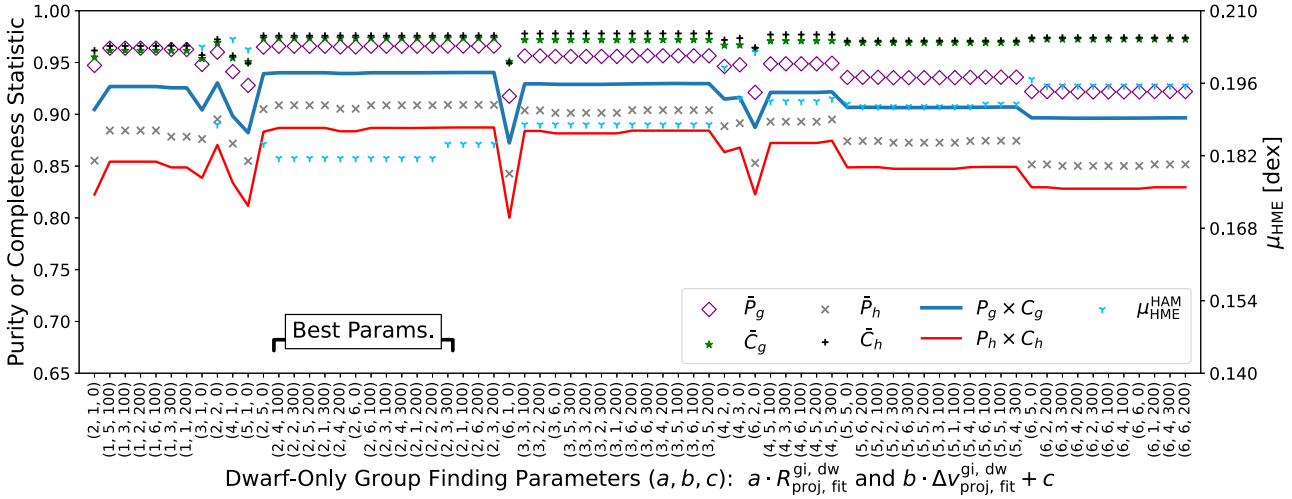


**Figure 11.** Scatter plot comparing group-finding performance for different parameter choices for the dwarf-only group-finding technique from Section 3.4. Each column corresponds to a different set of parameters for $R_{\rm proj,fit}^{\rm gi,dw}$ and $\Delta v_{\rm proj,fit}^{\rm gi,dw}$; e.g., (1, 2, 300) corresponds to $1R_{\rm proj,fit}^{\rm gi,dw}$ and $2\Delta$proj, fit$^{\rm gi,dw} + 300$ km s$^{-1}$ . Symbols and markers are as in Figure 10.

group center as a function of $N_{\rm giants}$, transitioning across the range $N_{\rm giants} = 4$–7. We run the group finder for all three algorithms, having fixed the multipliers to their optimal values set in Section 4.2. Despite the fact that our optimization used algorithm (i) above, we find that group finding is quite robust to the choice of group center definition. Changing from (i) to (ii) to (iii), we see a 1% then 3% drop in $\bar{C}_g$, coinciding with a 1% then 2% increase in $\bar{P}_g$. Because $\bar{P}_g$ is uniformly >98% in all three cases, we choose the simple average definition to maintain high $\bar{C}_g$ and to avoid potential systematic issues with a central-based definition.

### 4.4. Assessing Algorithm Performance

Now that the G3 group finder has been optimized, it is important to ensure that our algorithm is resilient to halo velocity bias, central galaxy location bias, and cosmic variance. In what follows, we show that the G3 group finder recovers highly pure and highly complete groups independent of these issues. We also assess its broader group-finding performance relative to other common algorithms applied at $z \sim 0$, including the RESOLVE and ECO FoF group catalogs.

#### 4.4.1. Purity and Completeness Distributions

We begin by testing whether purity and completeness distributions of groups recovered by the G3 algorithm are sensitive to halo velocity bias, central galaxy location bias, or cosmic variance. Figure 12 shows distributions of $P_g$ and $C_g$ in all 32 mocks, including markers for the mean values of $P_g$, $C_g$, $P_h$, and $C_h$, but distributions of $P_h$ and $C_h$ are omitted to maintain visual clarity. Our key finding is that we uniformly recover $\bar{P}_g$ and $\bar{P}_h$ of >88%, as well as $\bar{C}_g$ and $\bar{C}_h$ of >85%, independent of galaxy number density or halo bias (median values are ∼100% in all cases). We find only two systematic effects. First, there is a consistent drop in $\bar{C}_g$ in the highest-density mock (#8), which has twice the density of the entire SDSS and thus represents a substantial fluctuation in cosmic variance. Second, we find typically 3%–4% lower $\bar{P}_h$ and $\bar{C}_g$ in the central-offset case, which could be caused by an increase in cases of fragmentation during dwarf association. Such an effect is plausible because, in the central-offset mocks, central galaxies are not populated at halo centers, resulting in poorer group center estimation from Equation (3) for low-$N_{\rm giants}$ groups.

In all mocks, the purity and completeness statistics yielded by the G3 group finder are comparable with other recently published group finders. For example, Stothert et al. (2019) find typical purity of >88% and mean completeness >90% using their Markov clustering technique, and Rodriguez & Merchán (2020) find typical purity of >80% and mean completeness >90% by combining FoF with iterative halo-based algorithms.

Purity and completeness distributions are not necessarily uniform with halo mass, as a group finder could better recover smaller groups than larger groups or vice versa. Figure 13 shows distributions of purity and completeness as a function of halo mass for both the G3 group finder and the FoF group finder, applied to the fiducial mock catalogs. The FoF groups tend to attain higher $P_g$ and $P_h$ at all halo masses, while the G3 algorithm tends to attain higher $C_g$ and $C_h$ than FoF. The G3 algorithm also alleviates a trend toward lower completeness at high halo masses seen in FoF. Such halo mass dependence in purity and completeness is common among established group finders. For example, Duarte & Mamon (2014) show that the Berlind et al. (2006) linking lengths yield FoF groups that decrease in purity by ∼20% ("reliability" in their nomenclature) and decrease in completeness by ∼5% as group halo mass increases from $10^{12}\,M_\odot$ to $10^{15}\,M_\odot$. Similarly, Lim et al. (2017) show a ∼20% reduction in the fraction of SDSS groups with >95% completeness as halo mass increases from $10^{11}\,M_\odot$ to $10^{14.5}\,M_\odot$.

#### 4.4.2. Halo Mass Recovery

Finally, we use mocks to test whether our halo masses are robust to issues of halo velocity bias or central galaxy location bias. Figure 14 illustrates distributions of HAM halo mass errors as a function of $M_{\rm HAM}$, where the error is defined for each group as the median value of $\log M_{\rm HAM} - \log M_{\rm true}$ among group members. As seen in the upper four panels, the G3 algorithm yields a consistent distribution of halo mass errors in each of the fiducial, central-offset, $b_v = 0.8$, and $b_v = 1.2$ mocks, where each of these mocks combines the eight

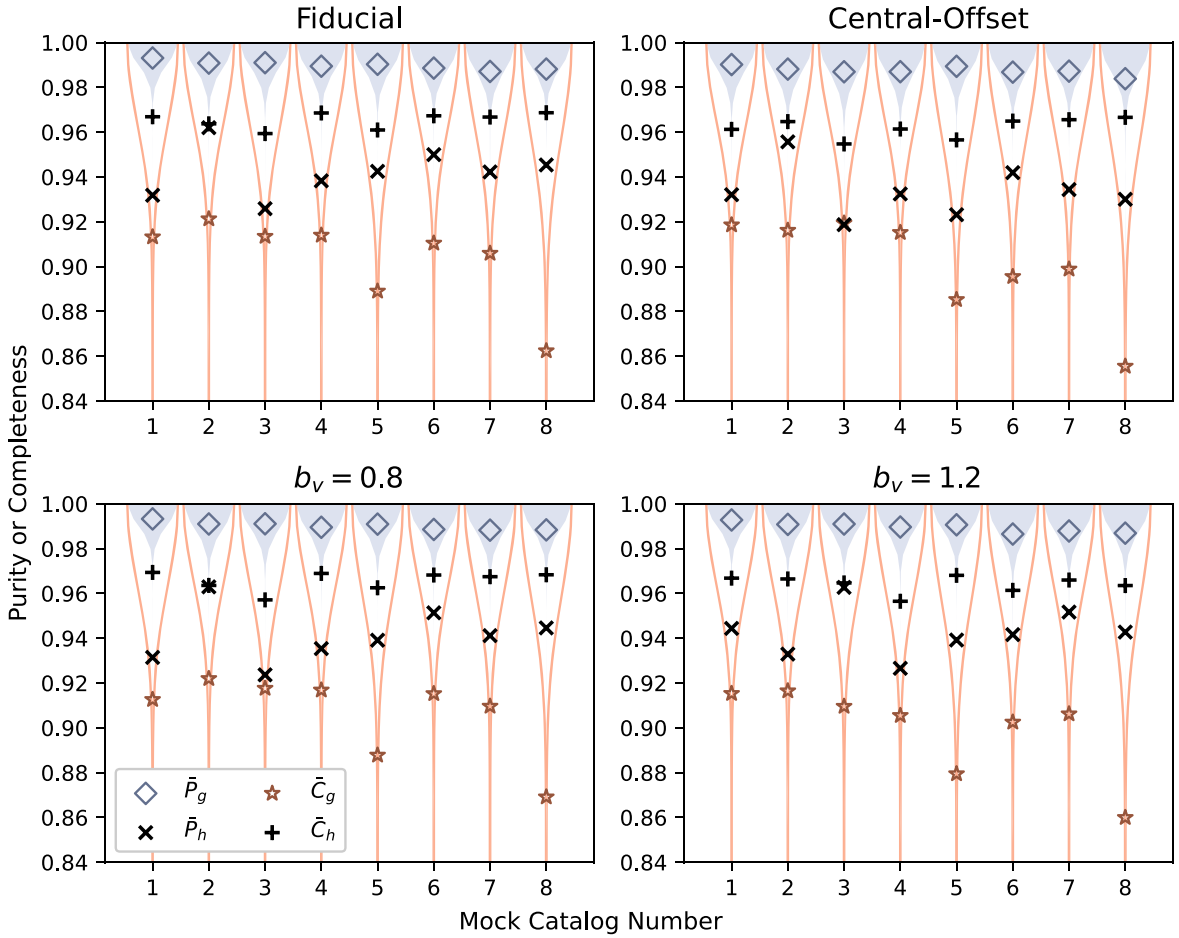


**Figure 12.** Violin plots illustrating distributions of group purity ($P_g$, blue) and group completeness ($C_g$, orange) for G3 groups in all 32 mocks, with additional black markers to illustrate mean $P_h$ and $C_h$ values. Violin widths are as in Figure 6. The four panels organize the mocks by fiducial, central-offset, $b_v = 0.8$, and $b_v = 1.2$ types, and the markers show the median and mean quantities. The mocks are labeled 1–8 in order of increasing galaxy number density, corresponding to 0.138, 0.158, 0.159, 0.183, 0.185, 0.223, 0.224, and 0.369 Mpc$^{-3}$. To improve clarity by ensuring that neighboring violins do not overlap, the horizontal positions of the violins are placed nonlinearly in number density.

separate mocks of that type. A small systematic offset of approximately +0.1 dex is observed in each case, likely resulting from variance of the mock mass functions, which reflect a slight underdensity on average compared with the Tinker et al. (2008) mass function that we use in HAM (e.g., see Figure 1). In an underdense mock, halos of a given mass are farther apart than they should be, such that the mean number density of these halos corresponds to higher halo mass on the Tinker et al. (2008) mass function, leading to systematically positive values of $M_{\rm HAM}$–$M_{\rm true}$. When we instead use the individual mock mass functions to assign HAM masses, we observe no systematic offset. We also note that, if we plot the average value of $\log M_{\rm HAM} - \log M_{\rm true}$ among group members, rather than the median, there is a steady rise to ∼0.4 dex at $\log M_{\rm HAM}[M_\odot] = 14.5$, which likely reflects increasing over-merging errors as halo mass increases. For comparison, the bottom panel in Figure 14 shows the equivalent distribution for the FoF group finder. The G3 algorithm provides a more accurate recovery of HAM masses for $\log M_{\rm HAM} = 11$–$13.5$, while FoF may perform slightly better in the high-mass regime.

## 5. The Group-integrated H I Mass–Halo Mass Relation

We now turn to using the RESOLVE and ECO G3 group catalogs to understand how group-integrated H I mass varies with group halo mass, i.e., the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation. The highly complete G3 group catalogs and comprehensive H I data contained within RESOLVE and ECO allow us to measure both the shape and scatter of the relation, the latter of which has not been addressed in previous observations. Analysis of the relation also serves as a useful test of our new group finder, by comparison to a conventional FoF group catalog, and of the updated ECO DR3 H I census, by comparison to RESOLVE's smaller but superior H I census.

### 5.1. The ECO DR3 Group H I–Halo Mass Relation

The left panel of Figure 15 shows the median $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation for RESOLVE and for ECO excluding RESOLVE-A ("ECO–RES-A"). We computed group-integrated H I masses $M_{\rm H\,I,grp}$ as the sum of all group members' preferred H I mass estimates, which are a combined data set of clean 21 cm detections, strong upper limits, deconfused observations, and estimates from photometric gas fractions (see Appendix A.6.2). Statistical errors on median values were determined using smoothed bootstrapping with 5000 resamples, and systematic errors were estimated from variations in median $M_{\rm H\,I,grp}$ due to cosmic variance. Specifically, cosmic variance errors for RESOLVE were estimated by dividing ECO into 10 RESOLVE-sized subvolumes. We have not attempted to

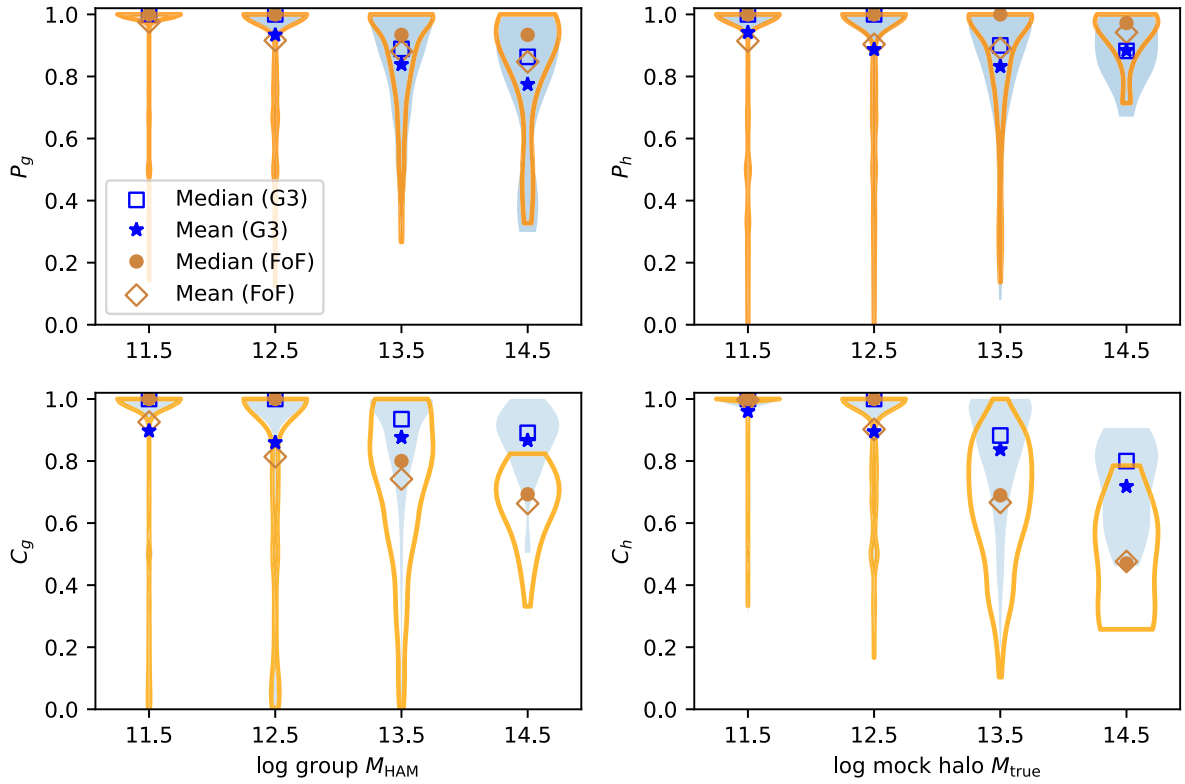


**Figure 13.** Violin plots illustrating distributions of purity and completeness as a function of halo mass for the G3 group finder (blue) and FoF (orange) for comparison. Both purity and completeness mappings are pictured (see Section 4.1). Violin widths are as in Figure 6. FoF (as implemented by our team with false-pair splitting) attains somewhat higher purity than the G3 algorithm, but G3 attains higher completeness, with dramatic improvement at high masses.

estimate cosmic variance errors for the larger ECO survey, which closely approximates the mean density of the entire SDSS (E16). We find that the median $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation is consistent for RESOLVE and ECO–RES-A given these errors, and we thus conclude that ECO's heavier reliance on photometric gas fractions does not bias the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation.

Figure 16 further tests our $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation by comparing the best-fit median relations derived with the G3 and FoF group catalog HAM masses. We use the halo model of Obuljen et al. (2019),

$$M_{\rm H\,I,grp}(M_{\rm halo}) = M_0\left(\frac{M_{\rm halo}}{M_{\rm min}}\right)^{\alpha}\exp\left(-\frac{M_{\rm min}}{M_{\rm halo}}\right), \quad (9)$$

which parameterizes $M_{\rm H\,I,grp}$ in terms of a normalization $M_0$, low-mass cutoff $M_{\rm min}$, and scaling parameter $\alpha$. We fit using the Levenberg–Marquadt least-squares fitting algorithm (Moré 1978) with weights corresponding to the statistical error for each $M_{\rm H\,I,grp}$ value. Table 2 lists the best-fitting model parameters for the models plotted in Figure 16, as well as for other sample definitions comparing G3 and FoF. We note that the best-fit values of $M_0$, $M_{\rm min}$, and $\alpha$ change only at the 1% level if we scale our $M_{\rm H\,I,grp}$ measurements by the boundary completeness correction factors from Section 2.4. Thus, these correction factors have no practical influence on our measurement of the relation. Figure 16 and Table 2 show that the FoF $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation is best fit by a smaller value of $\alpha$, reflecting lower group-integrated H I for massive FoF groups ($>10^{13.5}\,M_\odot$). This discrepancy may reflect the fact that FoF groups are substantially less complete on average than G3 groups in the high-mass regime (see lower panels of Figure 13).

### 5.2. The Shape and Scatter of the Relation in RESOLVE and ECO

The RESOLVE and ECO surveys enable robust measurements of both shape and scatter because they combine the highly complete, volume-limited G3 group catalogs with uniquely comprehensive H I data reaching down into the gas-poor galaxy regime. The left panel of Figure 15 shows the shape of the median $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation. It rises steeply at low halo masses, pivots in slope at a halo mass of $10^{11.4}$–$10^{11.5}\,M_\odot$, and then rises more gradually to high halo masses. The right panel of Figure 15 plots the group-integrated H I mass fraction, i.e., $M_{\rm H\,I,grp}/M_{\rm halo}$ as a function of halo mass alongside the group-integrated stellar and baryonic mass fractions. A parabolic fit to the ECO–RES-A $M_{\rm H\,I,grp}/M_{\rm halo}$ relation reveals a peak at $M_{\rm halo} = 11.4 \pm 0.1$, which was also seen by Guo et al. (2020). Via the central galaxy stellar mass–halo mass relation, the location of this peak coincides with the known transition in galaxy properties at a central stellar mass of $\sim 10^{9.7}\,M_\odot$ (see Dekel & Silk 1986; Garnett 2002; Dalcanton et al. 2004). As discussed by Kannappan et al. (2009, 2013), this mass is the "gas-richness threshold scale," i.e., the scale below which the central galaxy population is numerically dominated by gas-dominated dwarfs. In contrast, the group-integrated stellar and baryonic mass fractions peak around a halo mass of $\sim 10^{12.1}\,M_\odot$ in Figure 15, as also seen in previous work (Eckert et al. 2017; Guo et al. 2020). This peak corresponds to the "bimodality" scale observed by Kauffmann et al. (2003), which marks a transition in the relative number

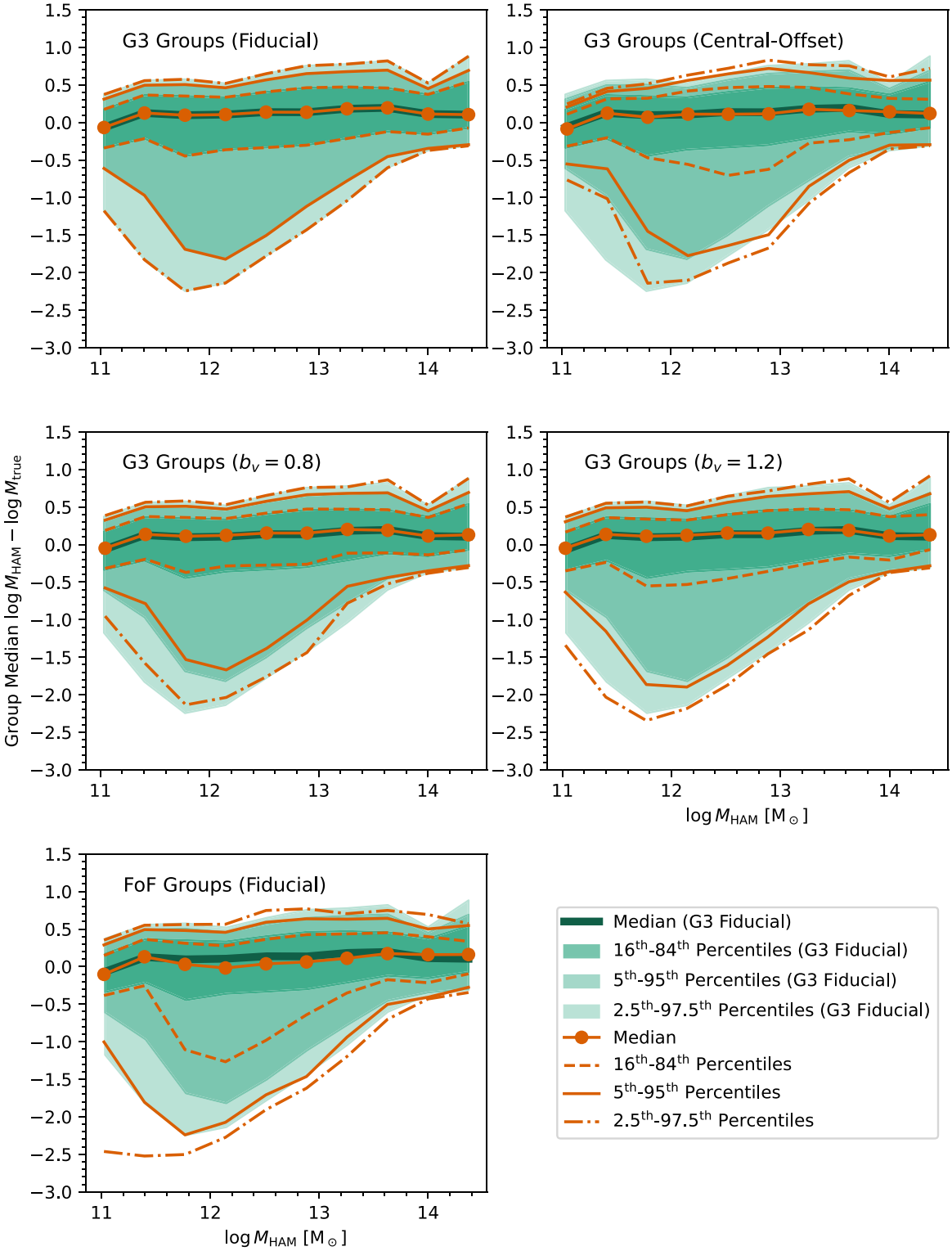


**Figure 14.** Conditional distributions of group median errors between "observed" HAM halo masses (group finding + HAM applied to mocks) and true halos masses (supplied with mocks), as a function of HAM mass. The upper four panels show the distributions for the G3 group finder applied to the fiducial, central-offset, and velocity bias mocks. To contrast with G3, the lower panel shows the distribution resulting from FoF group finding with false-pair splitting (E17). In each panel, the orange circles mark the median values at fixed $M_{\rm HAM}$ and dashed orange lines are drawn to enclose 68%, 90%, and 95% of the error distribution at fixed mass. For reference, the green shading in each panel shows the distribution from the G3 algorithm applied to fiducial mocks. These distributions show that the G3 algorithm consistently recovers HAM halo masses independent of halo velocity bias or central galaxy location bias.

dominance of disk-dominated galaxies with young stellar populations versus spheroid-dominated galaxies with old stellar populations. Kannappan et al. (2013) connected the threshold and bimodality scales to transitions in halo gas physics predicted in Dekel & Birnboim (2006; see also Birnboim & Dekel 2003; Kereš et al. 2005). In the simple, spherical model of Dekel & Birnboim (2006), as halos cross the threshold and bimodality halo mass scales, the halo gas shock heats outward

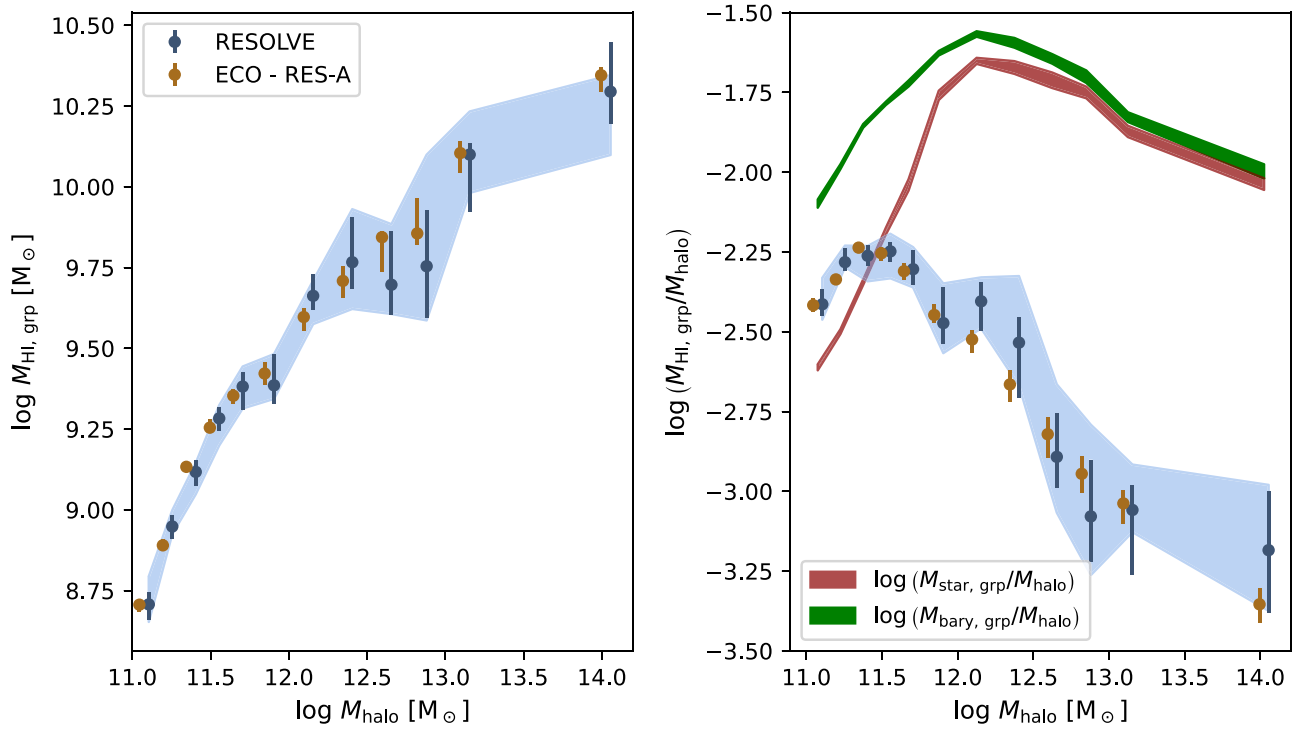


**Figure 15.** The group-integrated H I–halo mass relation in RESOLVE (blue) compared to ECO–RES-A (orange), both using G3 groups and HAM masses, illustrating the shape of the relation and its similarity between the RESOLVE and ECO–RES-A data sets, despite ECO's higher reliance on photometric gas fraction H I mass estimates. Left: group-integrated H I mass as a function of group halo mass for ECO–RES-A and RESOLVE. Error bars on the medians come from bootstrapping and the blue shading represents cosmic variance errors for RESOLVE. Right: group-integrated H I mass fractions as a function of group halo mass for RESOLVE and ECO–RES-A. The red and green lines are drawn for reference and represent the median group-integrated stellar and baryonic mass fractions in ECO–RES-A. The widths of the red and green lines represent statistical error from bootstrapping.

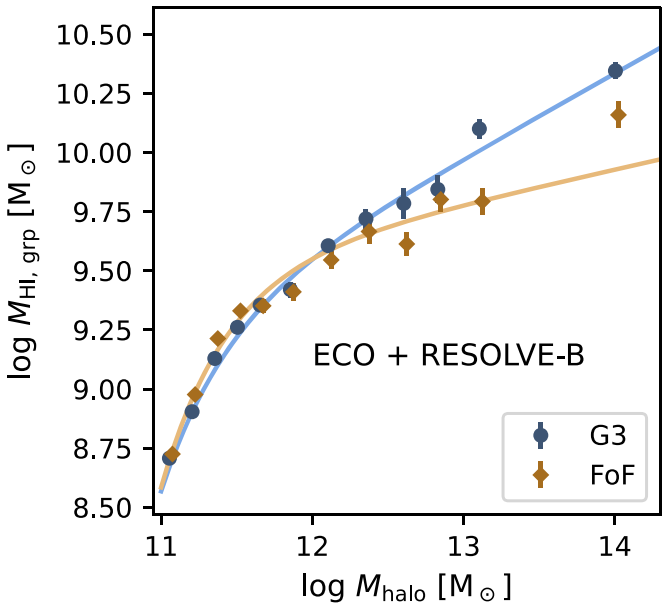


**Figure 16.** Model fits to the median ECO+RESOLVE-B $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation derived using the G3 compared to the FoF algorithms, illustrating inconsistency between these group finders. The best-fitting parameters of each model are listed in Table 2 along with the model parameters for other RESOLVE and ECO samples. Error bars come from bootstrapping.

to $0.1R_{\rm halo}$ and $1R_{\rm halo}$ respectively, thus limiting the rapid cold-mode accretion flows that efficiently fuel galaxies in lower-mass halos. Eckert et al. (2017) concluded that groups between the threshold and bimodality scales are transitioning from being gas-dominated to stellar-dominated, and the start of this transition seems to coincide with the pivot in slope we see in the median $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation.

Figure 17 shows the scatter in the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation. The typical median scatter in the relation is ∼0.3 dex and may depend on halo mass. A two-sample Kolmogorov–Smirnov (KS) test yields $>5\sigma$ confidence that the distribution of absolute residuals at $M_{\rm halo} < 10^{11.5} M_\odot$ is distinct from the distribution in the range $M_{\rm halo} = 10^{11.5\text{–}12.1} M_\odot$, but we do not find significant results for KS tests comparing higher halo mass bins where we also have fewer groups. In the well-sampled low halo mass regime, the scatter is not symmetric around the best-fitting model to the ECO–RES-A median relation. Overall, 94% of ECO–RES-A groups fall below +0.5 dex of the model, while only 82% land above −0.5 dex of the model. Consequently, the mean relation systematically underestimates the median relation, especially at low halo masses where the scatter is most asymmetric, albeit the absolute offset between the median and mean relation stays $\lesssim$0.2 dex. Physically, the scatter may reflect conditional dependence of $M_{\rm H\,I,grp}$ on secondary group metrics such as group $N_{\rm galaxies}$ (Guo et al. 2020), group crossing time (Ai & Zhu 2018), and large-scale environmental density (Stark et al. 2016, using RESOLVE).

Finally, we emphasize that our measured shape and scatter do not change significantly when we use the RESOLVE survey, which has a superior H I inventory with a much higher fraction of direct measurements. As seen in Figure 15, the median RESOLVE and ECO $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relations both show a pivot in slope and peak in $M_{\rm H\,I,grp}/M_{\rm halo}$ at the gas-richness threshold scale. As listed in Table 2, the median scatter

**Table 2**
Best-fitting Models to H I–Halo Mass Relation

| Group Catalog | $\log M_0$ ($\log M_\odot$) | $\log M_{\rm min}$ ($\log M_\odot$) | $\alpha$ | Scatter in $M_{\rm H\,I,grp}$ (dex) |
|---|---|---|---|---|
| ECO+RES-B —G3 | $9.329 \pm 1.000$ | $11.201 \pm 1.104$ | $0.359 \pm 0.023$ | 0.297 |
| ECO–RES-A —G3 | $9.328 \pm 0.957$ | $11.205 \pm 1.058$ | $0.362 \pm 0.024$ | 0.295 |
| Full RES —G3 | $9.385 \pm 0.786$ | $11.230 \pm 0.885$ | $0.313 \pm 0.045$ | 0.297 |
| Full ECO —G3 | $9.322 \pm 1.009$ | $11.191 \pm 1.108$ | $0.364 \pm 0.022$ | 0.294 |
| ECO+RES-B—FoF | $9.544 \pm 0.833$ | $11.323 \pm 0.957$ | $0.144 \pm 0.058$ | 0.305 |
| ECO–RES-A—FoF | $9.536 \pm 0.833$ | $11.319 \pm 0.958$ | $0.155 \pm 0.056$ | 0.304 |
| Full RES—FoF | $9.458 \pm 0.659$ | $11.228 \pm 0.725$ | $0.182 \pm 0.070$ | 0.317 |
| Full ECO—FoF | $9.536 \pm 0.850$ | $11.314 \pm 0.964$ | $0.156 \pm 0.055$ | 0.302 |

**Notes.** For comparison to G3 results, the bottom four rows present best-fitting model parameters using the FoF group finder. The scatter reported in the rightmost column is the median absolute deviation between the model predictions and underlying $M_{\rm H\,I,grp}$ data.

measurements for the RESOLVE and ECO $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relations are also completely consistent.

### 5.3. Comparison to Previous Work

We conclude by comparing our measurement of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation to previous observational and theoretical work, now using ECO+RES-B because this sample maximizes the number of groups available for analysis. The measurement of Guo et al. (2020), hereafter G20, is solely observational and thus most directly comparable to our work. As shown in Figure 18, the shape of the mean relation from G20 is in good agreement with the shape of the ECO+RES-B median relation for halo masses $\gtrsim 10^{11.5}\,M_\odot$ (the threshold scale). Going to lower halo masses, G20 measure increasingly higher $M_{\rm H\,I,grp}$ than found in ECO+RES-B, reaching a maximum discrepancy of 0.3 dex. The most obvious methodological difference between G20 and this work is their use of integration over entire halos and stacking in halo mass bins. G20 integrate the ALFALFA data cubes within $2r_{200}$ of each group center to extract a one-dimensional spectrum for each group, then stack the spectra for groups in halo mass bins. Comparing to their approach, our method of summing galactic H I within groups might underestimate the real $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation, as it does not probe contributions from intragroup H I or H I -rich galaxies below our surveys' luminosity completeness floors. Taking advantage of the fact that the full RESOLVE database includes additional galaxies below the luminosity completeness floor, we estimate a ∼5% increase (0.02 dex) in average H I mass per group for groups below $10^{11.5}\,M_\odot$ if we consider any associated galaxies below the luminosity floor to contribute as much gas as those above the floor. The increase is tiny because, although the RESOLVE database is >35% larger when adding galaxies below the luminosity floor, halos below the threshold mass generally contain solitary dwarf galaxies. Thus, if our discrepancy withG20 is astrophysical in origin, it likely reflects extended H I gas in dwarf galaxy halos out to $2r_{200}$. However, Roychowdhury et al. (2022) find that intragroup H I is substantial only for dwarf groups with 3–4 members, not solitary dwarfs like those typical of the $M_{\rm halo} < 10^{11.5}\,M_\odot$ mass regime.

Alternatively, the 0.3 dex discrepancy might be related to selection bias in the flux-limited Lim et al. (2017) SDSS group catalog used by G20, which is effectively volume-limited for galaxies with $M_r < -18.8$. G20 note that the group catalog is nearly complete above $M_{\rm halo} \sim 10^{11.5}\,M_\odot$, but they do not discuss the dependence of the incompleteness at lower mass on group properties. To resolve the observed discrepancy, the group catalog would have to be biased toward gas-rich halos below $M_{\rm halo} = 10^{11.5}\,M_\odot$.

Other studies have inferred the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation without direct H I observations. Obuljen et al. (2019), hereafter O19, constrained the relation by combining measurements of $M_{\rm H\,I,grp}$ for massive groups ($\gtrsim 10^{12.5}\,M_\odot$) with inferences from empirical modeling of H I clustering for lower-mass groups. Li et al. (2022), hereafter L22, inferred the shape of the relation using their own variant on photometric gas fraction H I estimation for all galaxies. Our measurement is roughly consistent with those of O19 and L22 below the threshold scale, but Figure 18 shows a discrepancy at higher halo masses. In particular, O19's constraints on $\log M_0$ and $\log M_{\rm min}$ are fully consistent with our best-fitting values in Table 2, but we fit a considerably smaller $\alpha$ value of ∼0.3 versus their $0.48 \pm 0.08$. This parameter controls the slope of the high-mass part of the relation and is thus sensitive to median $M_{\rm H\,I,grp}$ values for massive groups. As O19 determined the mean $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation by integrating the H I mass function in halo mass bins, they included inferred but not directly detected gas, and for that reason we would not expect to match their inferred relation in this mass regime. However, we do note that our best-fit value of $\alpha$ agrees with the relation they found using a clustering analysis of ALFALFA sources, which also agrees at high masses with the ALFALFA stacking analysis of G20. Also, Li et al. (2022) have offered a caveat to their own high $M_{\rm H\,I,grp}$ values at high $M_{\rm halo}$, suggesting their H I estimator could overpredict $M_{\rm H\,I,grp}$ if the gas fractions of satellite galaxies in massive groups are lower than those of calibrator galaxies with similar optical properties, as the estimator does not consider environment.

Figure 18 also shows three theoretical models of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation. In terms of shape, the simulation predictions of Calette et al. (2021), hereafter Ca21, best resemble our results, albeit they better match G20 at low halo masses and exceed both G20 and our measurements in the high-mass regime. In contrast, recent predictions from the semi-analytic models of Baugh et al. (2019), hereafter B19, building on the work of Kim et al. (2017) and Chauhan et al. (2020), hereafter C20, show a dramatic valley in the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation at $\log M_{\rm halo} = 12$–$12.5$ due to mechanical AGN feedback that sustains the hot halo and suspends the growth of $M_{\rm H\,I,grp}$. While C20 matches our G3 relation well at low halo masses, we do not detect the valley predicted by the B19 and C20 theoretical models, which could suggest that these models prescribe excess AGN feedback. However, Chauhan et al. (2021) showed that, when HAM and dynamical halo mass estimation techniques were

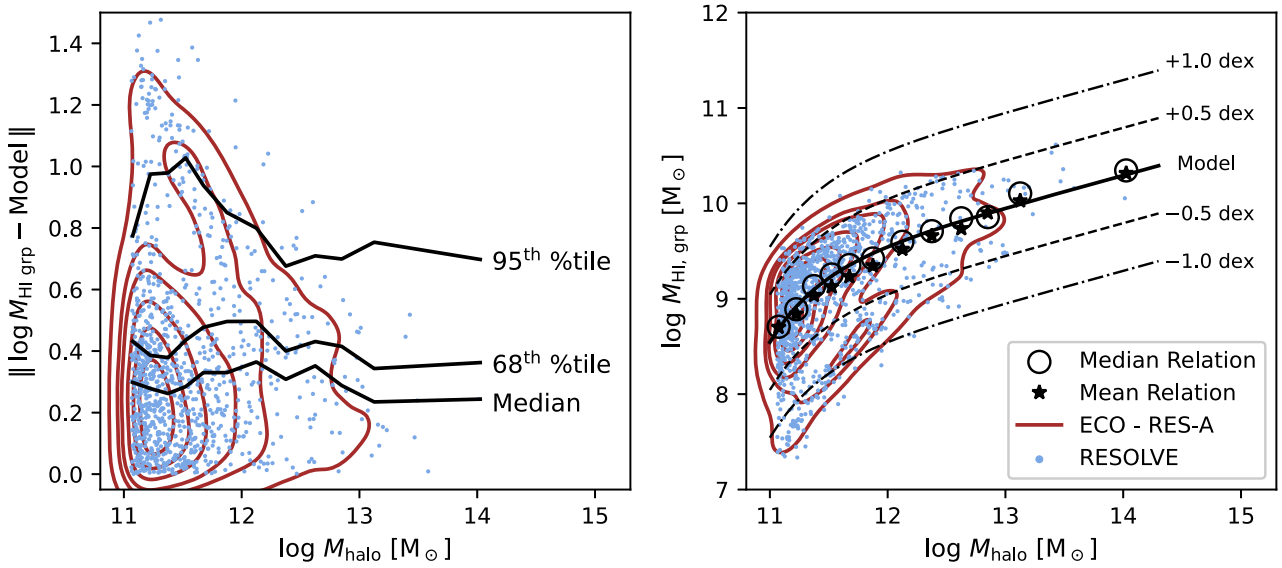


**Figure 17.** Scatter in the group-integrated H I–halo mass relation. Both panels use G3 groups and HAM masses. Left: scatter plot of absolute-valued residuals between individual $M_{\rm H\,I,grp}$ values and best-fitting median relations (Equation (9)) as in Table 2. RESOLVE is shown in blue and ECO–RES-A in red. Black lines denote the median, 68th percentile, and 9th percentile of the ECO–RES-A residuals as a function of halo mass. Right: scatter plot of $M_{\rm H\,I,grp}$ vs. $M_{\rm halo}$ for RESOLVE and ECO–RES-A, with the best-fitting model for ECO–RES-A overplotted. Dashed lines mark ±0.5 dex and ±1 dex offsets from the median ECO–RES-A relation, open circles show median $M_{\rm H\,I,grp}$ values for ECO–RES-A, and black stars show mean $M_{\rm H\,I,grp}$ values for ECO–RES-A.

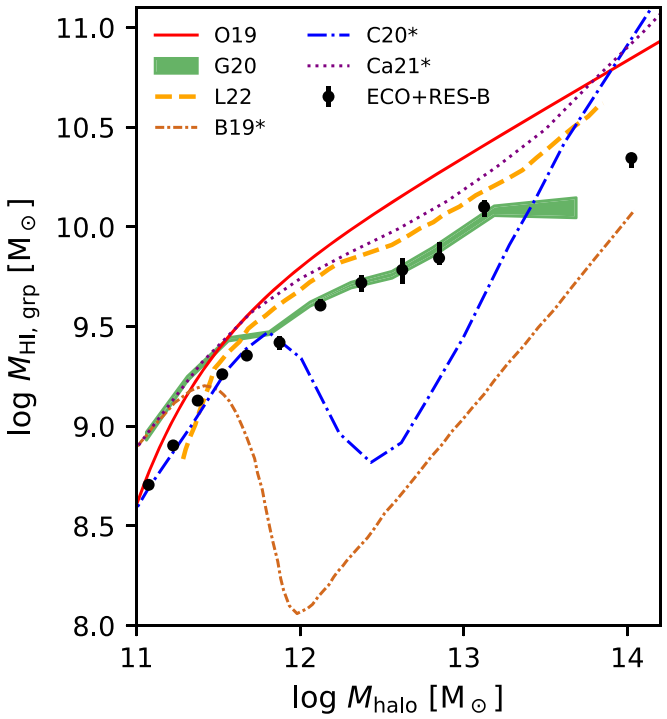


**Figure 18.** The group-integrated H I–halo mass relation from this study compared with several observational and theoretical results. The G3 median relation for ECO +RES-B is illustrated by the black data points. The observational mean relation and error of Guo et al. (2020) is shown in green shading, and inferred forms of the relation from Obuljen et al. (2019) and Li et al. (2022) are shown in solid red and dashed yellow, respectively. Theoretical predictions from Baugh et al. (2019), Chauhan et al. (2020), and Calette et al. (2021) are shown in dashed brown, dashed blue, and dotted purple, respectively.

applied in a mock catalog H I stacking analysis, the valley characteristic of the underlying model could not be recovered. The fact that neither we, O19, nor G20 observe such a valley may therefore be partially explained by observational analysis techniques.

Additionally, while B19, C20, and Ca21 predict halo-mass dependent scatter in $M_{\rm H\,I,grp}$, we find that none of these models reproduce the scatter we measure. B19 and C20 predict that the scatter's extent peaks at a value on the order of ∼1 dex in the valley region. Ca21 predict minimum scatter of 0.5 dex at $M_{\rm halo} = 10^{11}\,M_{\odot}$, rising to 1.2 dex by $M_{\rm halo} = 10^{12.5}\,M_{\odot}$. In contrast, our measured scatter is asymmetric, peaks at ∼1.3 dex at $\log M_{\rm halo} = 11.3$–$11.4$, and might weakly decrease with increasing halo mass, albeit our use of HAM may suppress the scatter's true extent. Comparing real and mock hydro simulation catalogs, E17 showed that the monotonic mapping underlying HAM suppresses scatter in group hot-to-cold gas ratios. Additionally, group-finding fragmentation and over-merging errors may alter the observed scatter. Quantifying the role of group-finding errors would require that we apply the G3 group finder and $M_{\rm H\,I,grp}$ measurement methodology to model-derived mock catalogs (as in E17 and Chauhan et al. 2021), which we have not attempted in this paper. Thus, quantifying the scatter in the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation and its cosmic evolution remains an open challenge.

## 6. Conclusion

In an era of deep H I surveys, unprecedented measurements of the cosmic evolution of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation are imminent. In this paper, we have prepared for future high-$z$ analysis by designing and optimizing a new galaxy group finder that can be consistently translated to incomplete, nonuniform, or flux-limited surveys such as LADUMA. We have applied this group finder to the highly complete and volume-limited $z \sim 0$ RESOLVE and ECO surveys and combined the resulting group catalogs with an expansion of the deep and uniquely comprehensive H I data available for these surveys. Together, the new group finder and expanded data have enabled us to measure both the shape and scatter of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation at $z \sim 0$ without stacking. We summarize our results as follows.

1. We designed a new, four-step group-finding algorithm and applied it to RESOLVE and ECO. To enable future comparisons with higher-$z$ data sets, we designed our group finder for translation to surveys with flux-limited, incomplete, or nonuniform data (Section 3, Figures 4–9).
2. Our new group-finding algorithm uses giant galaxies to define the initial basis for groups, mitigating the potential for bias if applied to a survey with an incomplete or nonuniform selection of dwarf galaxies.
3. Our new group-finding algorithm performs group membership refinement using relations calibrated on group $N_{\rm giants}$ and group-integrated luminosity. This approach avoids using halo quantities (e.g., virial radii) for membership refinement because these quantities are difficult to estimate for incomplete, nonuniform, or flux-limited surveys.
4. Using mock catalogs to test our group finder, we have shown that it yields consistent purity, completeness, and halo mass error statistics independent of cosmic variance, halo velocity bias, or central galaxy location bias. In comparison to friends-of-friends (FoF; implemented with false-pair splitting following E17), the G3 algorithm yields groups that are slightly less pure but substantially more complete, especially at high halo masses. G3 also improves on FoF in the accuracy of HAM halo mass estimates. These differences result in a steeper high-mass slope in the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation using G3 as opposed to FoF (Sections 4 and 5, Figures 12, 13, 14, and 16, and Table 2).
5. Our measurements of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation show a pivot at the gas-richness threshold scale of $M_{\rm halo} = 10^{11.4-11.5}\,M_{\odot}$, in agreement with the relations from O19, G20, and L22. This pivot corresponds to the peak in group-integrated H I mass fraction at the threshold mass, seen also in E17 and G20 (Section 5, Figures 15 and 18).
6. Below the threshold mass, our median $M_{\rm H\,I,grp}$ measurements are ∼0.3 dex lower than those of G20, possibly reflecting methodological differences. Our method considers only gas within galaxies in volume-limited surveys, whereas G20 use stacking to measure extended H I throughout galaxy halos in a flux-limited survey. In the low-halo-mass regime where we measure less H I than G20, our median $M_{\rm H\,I,grp}$ measurements agree well with indirect measurements of the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation (O19; L22). Conversely, above the threshold scale, our measurements are consistent with the relation of G20 but fall below those of O19 and L22 (Section 5, Figures 15 and 18).
7. We have presented the first observational measurement of the scatter in the $M_{\rm H\,I,grp}$–$M_{\rm halo}$ relation. The median scatter is on the order of 0.3 dex, but it is strongly asymmetric and might show weak halo-mass dependence. The theoretical models we considered predict larger scatter, perhaps in part due to our use of halo abundance matching, which E17 and Chauhan et al. (2021) have shown may suppress natural scatter (Section 5, Figure 17).

The group catalogs we have presented for the RESOLVE and ECO surveys define the parent group sample for the RESOLVE Gas in Galaxy Groups (G3) initiative. The goal of this initiative is to complement RESOLVE's stellar, atomic gas, and dynamical mass census with a complete inventory of multiphase gas in galaxy groups. This work lays the foundation for extending group multiphase gas analyses to next-generation higher-redshift surveys such as LADUMA.

## Acknowledgments

We are grateful to the anonymous referee, whose feedback has improved the quality of this work. We also thank Adrienne Erickcek, Andrew Mann, Mugdha Polimera, Matthew Bershady, Joshua Oppor, Jeremy Darling, Hayley Roberts, and Amir Kazemi-Moridani for valuable feedback at varying stages of the project. Z.L.H., S.J.K., and E.R.C. acknowledge support for this research from National Science Foundation (NSF) grant AST-1814486. Z.L.H. and D.S.C. are also supported through a North Carolina Space Grant Graduate Research Fellowship. S.J.K. and D.S.C. acknowledge support from NSF grant AST-2007351. A.J.B. acknowledges support from NSF grant AST-1814421. K.M.H. acknowledges financial support from the State Agency for Research of the Spanish Ministry of Science, Innovation and Universities through the "Center of Excellence Severo Ocho" awarded to the Instituto de Astrofísica de Andalucía (SEV-2017-0709), via participation in SKA-SPAIN, funded by the Ministry of Science and Innovation (MCIN), and financial support from grant RTI2018-096228-B-C31 (MCIU/AEI/FEDER,UE).

## Appendix
## ECO DR3 and RESOLVE DR4 Catalogs

As detailed in the following subsections, this work releases ECO DR3 and RESOLVE DR4, which include updates to catalog membership, redshifts, photometry, group catalogs, and the H I inventory. ECO DR3 and RESOLVE DR4, in addition to past RESOLVE and ECO data releases, are available online in searchable database format (https://resolve.astro.unc.edu/pages/database.php). Database users may consult our public tutorial (https://github.com/resolvesurvey/database-tutorial), which emphasizes sample definitions. Our luminosity- and volume-limited samples are defined by $M_r \leqslant -17.33$ in RESOLVE-A/ECO and $M_r \leqslant -17.0$ for RESOLVE-B, along with $4500 \leqslant cz_{\rm grp}[{\rm km\,s^{-1}}] \leqslant 7000$ for RESOLVE and $3000 \leqslant cz_{\rm grp}[{\rm km\,s^{-1}}] \leqslant 7000$ for ECO. The stellar mass and baryonic mass complete samples impose the same group $cz$ cuts, but are limited by (a) $\log M_* \geqslant 8.9$ in RESOLVE-A/ECO and $\log M_* \geqslant 8.7$ in RESOLVE-B for the stellar mass complete sample, or (b) $\log(M_* + 1.4M_{\rm H\,I}) \geqslant 9.3$ in RESOLVE-A/ECO and $\log(M_* + 1.4M_{\rm H\,I}) \geqslant 9.1$ in RESOLVE-B for the baryonic mass complete sample.

All physical quantities reported in these catalogs assume a ΛCDM cosmology with $H_0 = 70\ {\rm km\,s^{-1}}\,Mpc^{-1}$, $\Omega_{\rm m,0} = 0.3$, and $\Omega_{\Lambda,0} = 0.7$.

### A.1. Catalog Membership

In ECO DR3 and RESOLVE DR4, we refine our catalog membership as follows. Both surveys now exclude `ECO13860` (`rs1492` in RESOLVE-A), as recommended in the erratum to Hood et al. (2018). This target is too faint for RESOLVE membership and has faulty photometry due to overlap with a larger galaxy. Additionally, we reject `ECO13245`, a star that was misclassified as a galaxy in the original ECO catalog. No other such misclassifications were identified, but we do note that the existing photometry of `ECO11472` is centered on a superposed star, so its coordinates are inaccurate. In RESOLVE, we add `rs1519`, `rs1520`, `rs1521`, and `rs1522`. These galaxies are

**Table 3**
Duplicate Entries Rejected in ECO DR3

| ECO DR2 Identifier[a] | ECO DR3 Identifier[b] | Classification[c] |
|---|---|---|
| ECO01050 | ECO00059 | 2 |
| ECO05088 | ECO11318 | 2 |
| ECO06326 | ECO04128 | 2 |
| ECO06598 | ECO13581 | 1 |
| ECO06655 | ECO13599 | 1 |
| ECO10553 | ECO03588 | 2 |
| ECO11597 | ECO04983 | 1 |
| ECO12027 | ECO03434 | 2 |
| ECO12274 | ECO00805 | 2 |
| ECO12301 | ECO12301 | 2 |
| ECO12315 | ECO08456 | 2 |
| ECO12604 | ECO12602 | 1 |
| ECO13577 | ECO03705 | 1 |
| ECO13580 | ECO06454 | 1 |
| ECO13598 | ECO03650 | 1 |

**Notes.**
[a] Identifier of ECO DR2 database entry now rejected from ECO DR3.
[b] Identifier of matched entry in ECO DR3.
[c] If "1," the two entries were exact copies; if "2," the rejected entry was centered on a feature of the galaxy (e.g., a spiral arm).

located in the overlap of RESOLVE-A and ECO, but were inadvertently missed in previous RESOLVE data releases, despite having official ECO names. In RESOLVE DR4, we inherit redshifts, photometry, H I data, and environment metrics for these galaxies from ECO DR3.

### *A.2. Duplicate ECO Database Entries*

ECO DR2 inadvertently included several duplicate entries, including a few exact carbon-copy galaxies. To perform a thorough check for duplicates, we cross-matched ECO with itself, using each galaxy's effective radius as an adaptive matching radius (typically 5″–6″). The resulting candidates were then inspected in the SDSS DR12 ImageList Tool and classified visually. Only 15 of these candidates (0.1% of ECO DR2's 13,878 galaxies) were flagged as genuine duplicates. Table 3 lists these duplicates, their survey identifiers from ECO DR2, and the reasons they were flagged as duplicates. All ECO DR3 tables presented in the following subsections, as well as the analysis of this paper, exclude database entries classified as duplicates. However, Table 4 does include a visual inspection flag `vif` for nonduplicate entries created as part of the process of inspecting potential duplicates, which can be used to identify some merging systems or close galaxy pairs. Possible values are as follows: (0) the galaxy is assumed to not be a duplicate, but has not been visually inspected; (−1) the galaxy is the primary target within a merger remnant representing one catalog object; (−2) the galaxy was duplicated in ECO DR2 and only this entry is kept in ECO DR3; (−3) ECO DR2 included a second entry for this galaxy that was centered on a spiral arm or other feature; (−4) the galaxy has a nearby object at similar redshift, but no obvious connecting material or peculiar morphological features indicate a merger in progress; (−5) the galaxy is within the effective radius of a second galaxy or vice versa; (−6) the galaxy is undergoing a merger, but both cores are visually identifiable.

**Table 4**
ECO DR3 Photometry and SED Modeling Outputs

| Column | Designation | Description |
|---|---|---|
| 1 | `name` | ECO Galaxy Identifier |
| 2 | `resname` | RESOLVE Galaxy Identifier |
| 3 | `radeg` | R.A. J2000 |
| 4 | `dedeg` | Decl. J2000 |
| 5 | `cz` | Local Group-corrected Recession Velocity of Galaxy |
| 6 | `cze16` | Local Group-corrected Recession Velocity of Galaxy from E16 |
| 7 | `czhel` | Heliocentric Recession Velocity of Galaxy |
| 8 | `loscmvgdist` | Line-of-sight Comoving Distance to Galaxy |
| 9 | `vif` | Visual Inspection Flag (1) |
| 10 | `fm15` | ECO DR1 Moffett+15 Membership Flag |
| 11 | `nuvmag` | Apparent Magnitude in GALEX NUV Band (2) |
| 12 | `e_nuvmag` | Error in GALEX NUV Band |
| 13 | `umag` | Apparent Magnitude in SDSS *u* Band (2) |
| 14 | `e_umag` | Error in SDSS *u* Band |
| 15 | `gmag` | Apparent Magnitude in SDSS *g* Band (2) |
| 16 | `e_gmag` | Error in SDSS *g* Band |
| 17 | `rmag` | Apparent Magnitude in SDSS *r* Band (2) |
| 18 | `e_rmag` | Error in SDSS *r* Band |
| 19 | `imag` | Apparent Magnitude in SDSS *i* Band (2) |
| 20 | `e_imag` | Error in SDSS *i* Band |
| 21 | `zmag` | Apparent Magnitude in SDSS *z* Band (2) |
| 22 | `e_zmag` | Error in SDSS *z* Band |
| 23 | `2jmag` | Apparent Magnitude in 2MASS *J* Band (2) |
| 24 | `e_2jmag` | Error in 2MASS *J* Band |
| 25 | `2hmag` | Apparent Magnitude in 2MASS *H* Band (2) |
| 26 | `e_2hmag` | Error in 2MASS *H* Band |
| 27 | `2kmag` | Apparent Magnitude in 2MASS *K* Band (2) |
| 28 | `e_2kmag` | Error in 2MASS *K* Band |
| 29 | `uymag` | Apparent Magnitude in UKIDSS *Y* Band (2) |
| 30 | `e_uymag` | Error in UKIDSS *Y* Band |
| 31 | `uhmag` | Apparent Magnitude in UKIDSS *H* Band (2) |
| 32 | `e_uhmag` | Error in UKIDSS *H* Band |
| 33 | `ukmag` | Apparent Magnitude in UKIDSS *K* Band (2) |
| 34 | `e_ukmag` | Error in UKIDSS *K* Band |
| 35 | `extnuv` | Foreground Extinction in GALEX NUV Band |
| 36 | `extu` | Foreground Extinction in SDSS *u* Band |
| 37 | `extg` | Foreground Extinction in SDSS *g* Band |
| 38 | `extr` | Foreground Extinction in SDSS *r* Band |
| 39 | `exti` | Foreground Extinction in SDSS *i* Band |
| 40 | `extz` | Foreground Extinction in SDSS *z* Band |
| 41 | `exty` | Foreground Extinction in UKIDSS *Y* Band |
| 42 | `extj` | Foreground Extinction in 2MASS *J* Band |
| 43 | `exth` | Foreground Extinction in 2MASS and UKIDSS *H* Bands |
| 44 | `extk` | Foreground Extinction in 2MASS and UKIDSS *K* Bands |
| 45 | `rejectedphot` | Photometry Rejected During SED Fitting |
| 46 | `badrphot` | Bad *r*-band Photometry Flag (3) |
| 47 | `logmstar` | Log Galaxy Stellar Mass (4) |
| 48 | `absrmag` | Absolute Magnitude in SDSS *r* Band (5) |
| 49 | `absrmage16` | Absolute Magnitude in SDSS *r* Band from E16(5) |

**Table 4**
(Continued)

| Column | Designation | Description |
|---|---|---|
| 50 | modelabsrmag | Rest Frame SED Modeled Absolute Magnitude in SDSS $r$ Band (6) |
| 51 | modelnuvr | Rest Frame SED Modeled (NUV $-r$) Color (6) |
| 52 | modelur | Rest Frame SED Modeled ($u-r$) Color (6) |
| 53 | modelui | Rest Frame SED Modeled ($u-i$) Color (6) |
| 54 | modeluj | Rest Frame SED Modeled ($u-J$) Color (6) |
| 55 | modeluk | Rest Frame SED Modeled ($u-K$) Color (6) |
| 56 | modelgr | Rest Frame SED Modeled ($g-r$) Color (6) |
| 57 | modelgi | Rest Frame SED Modeled ($g-i$) Color (6) |
| 58 | modelgj | Rest Frame SED Modeled ($g-J$) Color (6) |
| 59 | modelgk | Rest Frame SED Modeled ($g-K$) Color (6) |
| 60 | modelurcorr | Rest Frame SED Modeled ($u-r$) Color Corrected for Internal Extinction (7) |
| 61 | r50 | Half-light Radius in $r$ Band (8) |
| 62 | r90 | 90%-light Radius in $r$ Band (8) |
| 63 | axialratio | Axial Ratio |
| 64 | mudelta | Morphological metric $\mu_\Delta$ (9) |

**Notes.** Only a portion of this table is shown here to demonstrate its form and content. The full table is available online in machine-readable format. (1) Inspection flag resulting from duplicate search. No duplicate database entries are included in this table, but this data column tabulates the reasons why particular galaxies were inspected or eliminated as possible duplicates. Possible values are described in the machine-readable table and in Appendix A.2. (2) Custom remeasured value from E16, provided without foreground extinction correction. (3) Bad $r$-band photometry flag (see Appendix A.4), marking galaxies for which effective radii were previously set to arbitrary or negative values due to pipeline failures. 0 = No $r$-band photometry issue, 1=bad but usable $r$-band photometry or absrmag inferred from stellar mass, 2=unreliable $r$-band photometry and stellar mass. (4) Log galaxy stellar mass derived using cz to calculate cosmological luminosity distances. Please refer to badrphot when using this column. (5) Absolute magnitude based on custom $r$-band photometry of E16, foreground extinction correction included, scaled to cz or cze16 as discussed in Appendix A.3. Please refer to badrphot when using these columns. (6) Values include foreground extinction corrections and k-corrections, but not internal extinction corrections. (7) As explained in Appendix A.4, we expect nonuniformity in the internal extinction corrections for ECO outside RESOLVE. We recommend modelur for ECO analyses requiring uniformity. (8) Values are set to –999 where badrphot>0. (9) Morphological metric $\mu_\Delta$ defined in Kannappan et al. (2013), derived using cz to calculate cosmological angular diameter distances. Values are set to –999 where badrphot>0.

(This table is available in its entirety in machine-readable form.)

### A.3. Updated ECO Redshifts

With ECO DR3, we include updated redshifts to ensure consistency with RESOLVE. For galaxies belonging to both RESOLVE and ECO, the ECO DR3 redshifts are drawn directly from RESOLVE so that the RESOLVE and ECO redshifts match identically for shared galaxies. RESOLVE DR4 does not include any changes to the redshifts of RESOLVE galaxies. For galaxies belonging only to ECO, we update the original redshifts (compiled from SDSS, HyperLEDA, etc.; see Moffett et al. 2015) where possible with redshifts from SDSS DR7 and DR8. Redshifts from SDSS DR8 were obtained from the Max Planck Institute for Astrophysics–Johns Hopkins University (MPA-JHU) line measurement catalog, ignoring entries for which the RELEASE keyword was set to "special," "extra," or "extraspecial," which represent duplicate or repeat plates of the primary survey targets (Adelman-McCarthy et al. 2007), and ignoring entries for which the RELIABLE or zWarning flags indicated unreliable redshifts. Likewise, SDSS DR7 redshifts were obtained using the online CasJobs/SciServer website by selecting galaxies for which $z>0$, zWarning >0, and zConfFinal >0.35 (following Strauss et al. 2002; Malavasi et al. 2020). We used the TOPCAT software (Taylor 2017) to perform a 6" optical cross-match in order to identify galaxies in both ECO and SDSS DR7/DR8. In cases where an ECO galaxy had a match in both SDSS DR7 and DR8, we chose the DR7 match, as SDSS DR7 (not DR8) was the primary contributor of redshifts to RESOLVE (Eckert et al. 2015). As provided in previous RESOLVE and ECO data releases, all redshifts obtained from the SDSS were converted from the heliocentric reference frame to the Local Group reference frame using

$$v_{\rm LG} = v_{\rm hel} + (300\ {\rm km\ s^{-1}})\cos(b)\sin(l), \quad \text{(A1)}$$

where $v_{\rm hel}$ is the heliocentric recessional velocity, $b$ is the Galactic latitude, and $l$ is the Galactic longitude.

Although these updated redshifts (cz) are the "primary" redshift measurements for ECO DR3, Table 4 also tabulates the galaxy redshifts as used by E16 and E17 (cze16). This column enables the reconstruction of physical quantities consistent with the E16 and E17 analyses, as we report physical quantities based on angular diameter or luminosity distances calculated from cz using ΛCDM (comoving distances are provided in the table), rather than based on Hubble Law distances calculated from cze16 (as in E16/E17). Additionally, Table 5 includes our group center redshifts for the G3 and updated FoF algorithms (see Appendix A.5). Users may wish to rescale our physical quantities using these group redshifts to mitigate peculiar velocity distortions, but our tabulated comoving distances are based on individual galaxy redshifts.

### A.4. Photometry

We provide improved photometric quality information and derived photometric parameters with RESOLVE DR4 and ECO DR3, reflecting better treatment of galaxies with unreliable filter data and/or radii. We have not remeasured any RESOLVE and ECO photometry used by E15/E16/E17, except for rs0106, which was reprocessed by K. Eckert using the code from E15. We have, however, indicated poor measurements and in some cases rejected them from calculations. Table 4 contains filter photometry, stellar masses, colors output from SED modeling, and morphological information for ECO. Analogous quantities for RESOLVE are in Table 6. These tables include recomputed SED modeling outputs using the code of Kannappan et al. (2013) as implemented in E16/E17, but now excluding input filter magnitudes that are unreliable. In particular, we exclude magnitudes if their errors are extremely large (i.e., $\sigma_{\rm NUV}>3.6$ mag, $\sigma_u>3$ mag, $\sigma_{griz}>1.7$ mag, or $\sigma_{YJHK}>2.3$ mag). Additionally, we reject outliers in (NUV $-u$) versus ($u-r$) color space, rejecting NUV when (NUV $-u$) $<-4$ and $-2.5>(u-r)<4.5$, and

**Table 5**
ECO DR3 Galaxy Groups

| Column | Designation | Description |
|---|---|---|
| 1 | name | ECO Galaxy Identifier |
| 2 | involve17 | In-volume Flag for E17 Groups (1) |
| 3 | g3grp | G3 Group Identifier |
| 4 | g3grpradeg | G3 Group Center R.A. J2000 |
| 5 | g3grpdedeg | G3 Group Center Decl. J2000 |
| 6 | g3grpcz | G3 Group Center *cz* (2) |
| 7 | involg3 | In-volume Flag for G3 Groups (1) |
| 8 | g3grpngi | Number of Giants in G3 Group |
| 9 | g3grpndw | Number of Dwarfs in G3 Group |
| 10 | g3grpabsrmag | G3 Group-integrated *r*-Band Absolute Magnitude |
| 11 | g3grpgiantabsrmag | G3 Group-integrated Giant *r*-Band Absolute Magnitude |
| 12 | g3logmhvir | G3 HAM Log Group Mass (337b) (3) |
| 13 | g3logmh200 | G3 HAM Log Group Mass (200b) (4) |
| 14 | g3grpmhi | G3 Log Group-integrated H I Mass |
| 15 | g3fc | G3 Central Galaxy Flag (5) |
| 16 | ccbremapped | Boundary Completeness Correction Factor (6) |
| 17 | fofgrp | FoF Group Identifier |
| 18 | fofgrpradeg | FoF Group Center R.A. J2000 |
| 19 | fofgrpdedeg | FoF Group Center Decl. J2000 |
| 20 | fofgrpcz | FoF Group Center *cz* (2) |
| 21 | involfof | In-volume Flag for FoF Groups (1) |
| 22 | fofgrpn | Number of Galaxies in FoF Group |
| 23 | fofgrpabsrmag | FoF Group-integrated *r*-Band Absolute Magnitude |
| 24 | foflogmhvir | FoF HAM Log Group Mass (337b) (3) |
| 25 | foflogmh200 | FoF HAM Log Group Mass (200b) (4) |
| 26 | fofgrpmhi | FoF Log Group-integrated H I Mass |
| 27 | foffc | FoF Central Galaxy Flag (5) |

**Notes.** Only a portion of this table is shown here to demonstrate its form and content. The full table is available online in machine-readable format. (1) Group in-volume flag, set to 1 if group satisfies $3000 < cz_{\rm grp}[{\rm km\ s^{-1}}] < 7000$. involve17 is calculated using Grpcz from Table 1 of E17, included here to enable reconstruction of prior survey membership by joining to absrmagE16. (2) Local Group-corrected recession velocity of group center, derived using cz in Table 4. (3) HAM performed using the Tinker et al. (2008) halo mass function. (4) HAM performed using the Warren et al. (2006) halo mass function. (5) Set to 1 if galaxy is the group's central (most luminous in *r* band), or 0 otherwise. (6) Correction factors for group galaxy number count incompleteness due to group peculiar velocities extending beyond the RESOLVE and ECO redshift limits; affects only Coma and two massive groups. If >1, the group extends beyond ECO's redshift range. Values as computed by E16 based on cze16, mapped to our new G3 group catalog as described in Appendix A.5.

(This table is available in its entirety in machine-readable form.)

rejecting $u$ when (NUV $-u$) $< -4$ but $(u-r)$ is outside $-2.5$–$4.5$ mag. All rejected magnitudes are still reported in Tables 4 and 6, to ensure reproducibility of E15/E16/E17, but they are recorded in the rejectedphot columns.

Additionally, we have generated a flag badrphot to mark 113 ECO galaxies for which *r*-band 50% and 90% light radii were previously set to arbitrary or negative values due to pipeline failures in the presence of nearby bright objects or SDSS imaging defects. Based on analysis of photometric correlations, we found that 85 of these galaxies still had usable stellar masses and absolute *r*-band magnitudes, deviating from plausible values by less than ∼0.5 dex or ∼1 mag. We marked these cases with badrphot=1. For the remaining 28, we attempted to validate their stellar masses via visual inspection and analysis of gas masses and H I linewidths. If the stellar mass was usable, we set badrphot=1 and inferred an estimate of $M_r$ from a forward fit to the $M_r$ versus $M_*$ relation for RESOLVE ($M_r = -1.82[\log M_*] - 1.915$), since both parameters are necessary for survey definition. For the final 22 cases in which we deemed both stellar mass and $M_r$ unusable, we have tabulated nominal (but highly unreliable) $M_*$ and $M_r$ values and set the flag badrphot=2. For completeness of the ECO database, we have still performed group finding including these galaxies with the nominal values. For all 113 galaxies, radii and $\mu_\Delta$ are set to –999.

Finally, we note an important limitation of the ECO database affecting NUV magnitudes and internal extinction corrections. ECO, unlike RESOLVE, does not have full deep GALEX NUV coverage. E16 required that the NUV exposure be >1000 s to measure an NUV magnitude for the ECO database, and roughly half of ECO galaxies lack GALEX NUV data at this depth. As our stellar population synthesis modeling uses the E16 custom NUV magnitudes by default, internal extinction corrections computed for galaxies without such NUV magnitudes are not consistent with corrections computed for galaxies with them (see Figure 19); thus, we currently recommend photometry without internal extinction corrections (e.g., modelur as opposed to modelurcorr) for ECO analyses requiring uniformity. We hope to address the nonuniformity of ECO internal extinction corrections in a future data release.

### A.5. Group Catalogs

Two group catalogs are available in RESOLVE DR4 and ECO DR3. Foremost, we release the group catalog for the Gas in Galaxy Groups initiative, created by the algorithm described in Section 3. We also provide the updated friends-of-friends (FoF) catalogs that have been used in this paper (referred to as "FoF" as opposed to "E17 FoF"). Following Eckert et al. (2017), the FoF group catalog was created by applying FoF using linking constants of $b_\perp = 0.07$ and $b_{||} = 1.1$ with an additional procedure to split false pairs. However, the updated FoF algorithm has been improved by performing linking in comoving coordinates, using ECO's mean comoving number density ($0.067h^3$ Mpc$^{-3}$) to compute the final linking lengths. We also corrected a minor issue that caused five groups to be incorrectly flagged as false FoF pairs and thus incorrectly labeled as $N_{\rm galaxies} = 1$ groups in Eckert et al. (2017). For ECO specifically, the updated FoF group catalogs also reflect the duplicate entry rejections (Appendix A.2) and updated redshifts (Appendix A.3).

For dwarf galaxies in the RESOLVE and ECO databases that are below the survey luminosity completeness floors—and therefore excluded during group finding—we have assigned group membership after group finding by association to previously identified groups. These group membership assignments are strictly for database completeness and are not used to compute group metrics for our Section 5 analysis. Subfloor dwarfs for G3 groups are associated using the association methods described in Section 3, but for our FoF groups, we use an association algorithm updating E17. E17 associated subfloor dwarfs to a given FoF group within (a) the theoretical virial radius inferred from the assigned halo mass and (b) the larger of the line-of-sight linking length or three times the observed

**Table 6**
RESOLVE DR4

| Column | Designation | Description |
|---|---|---|
| 1 | name | RESOLVE Galaxy Identifier |
| 2 | radeg | R.A. J2000 |
| 3 | dedeg | Decl. J2000 |
| 4 | cz | Local Group-corrected Recession Velocity of Galaxy (1) |
| 5 | czhel | Heliocentric Recession Velocity of Galaxy |
| 6 | loscmvgdist | Line-of-sight Comoving Distance to Galaxy |
| 7 | nuvmag | Apparent Magnitude in GALEX NUV Band (2) |
| 8 | e_nuvmag | Error in GALEX NUV Band |
| 9 | umag | Apparent Magnitude in SDSS $u$ Band (2) |
| 10 | e_umag | Error in SDSS $u$ Band |
| 11 | gmag | Apparent Magnitude in SDSS $g$ Band (2) |
| 12 | e_gmag | Error in SDSS $g$ Band |
| 13 | rmag | Apparent Magnitude in SDSS $r$ Band (2) |
| 14 | e_rmag | Error in SDSS $r$ Band |
| 15 | imag | Apparent Magnitude in SDSS $i$ Band (2) |
| 16 | e_imag | Error in SDSS $i$ Band |
| 17 | zmag | Apparent Magnitude in SDSS $z$ Band (2) |
| 18 | e_zmag | Error in SDSS $z$ Band |
| 19 | 2jmag | Apparent Magnitude in 2MASS $J$ Band (2) |
| 20 | e_2jmag | Error in 2MASS $J$ Band |
| 21 | 2hmag | Apparent Magnitude in 2MASS $H$ Band (2) |
| 22 | e_2hmag | Error in 2MASS $H$ Band |
| 23 | 2kmag | Apparent Magnitude in 2MASS $K$ Band (2) |
| 24 | e_2kmag | Error in 2MASS $K$ Band |
| 25 | uymag | Apparent Magnitude in UKIDSS $Y$ Band (2) |
| 26 | e_uymag | Error in UKIDSS $Y$ Band |
| 27 | uhmag | Apparent Magnitude in UKIDSS $H$ Band (2) |
| 28 | e_uhmag | Error in UKIDSS $H$ Band |
| 29 | ukmag | Apparent Magnitude in UKIDSS $K$ Band (2) |
| 30 | e_ukmag | Error in UKIDSS $K$ Band |
| 31 | extnuv | Foreground Extinction in GALEX NUV Band (3) |
| 32 | extu | Foreground Extinction in SDSS $u$ Band |
| 33 | extg | Foreground Extinction in SDSS $g$ Band |
| 34 | extr | Foreground Extinction in SDSS $r$ Band |
| 35 | exti | Foreground Extinction in SDSS $i$ Band |
| 36 | extz | Foreground Extinction in SDSS $z$ Band |
| 37 | exty | Foreground Extinction in UKIDSS $Y$ Band |
| 38 | extj | Foreground Extinction in 2MASS $J$ Band |
| 39 | exth | Foreground Extinction in 2MASS and UKIDSS $H$ Bands |
| 40 | extk | Foreground Extinction in 2MASS and UKIDSS $K$ Bands |
| 41 | rejectedphot | Photometry Rejected During SED Fitting |
| 42 | logmstar | Log Galaxy Stellar Mass (4) |
| 43 | absrmag | Absolute Magnitude in SDSS $r$ Band (5) |
| 44 | modelabsrmag | Rest Frame SED Modeled Absolute Magnitude in SDSS $r$ Band (6) |
| 45 | modelnuvr | Rest Frame SED Modeled (NUV $-r$) Color (6) |
| 46 | modelur | Rest Frame SED Modeled ($u-r$) Color (6) |
| 47 | modelui | Rest Frame SED Modeled($u-i$) Color (6) |
| 48 | modeluj | Rest Frame SED Modeled ($u-J$) Color (6) |
| 49 | modeluk | Rest Frame SED Modeled ($u-K$) Color (6) |
| 50 | modelgr | Rest Frame SED Modeled ($g-r$) Color (6) |
| 51 | modelgi | Rest Frame SED Modeled ($g-i$) Color (6) |
| 52 | modelgj | Rest Frame SED Modeled ($g-J$) Color (6) |
| 53 | modelgk | Rest Frame SED Modeled ($g-K$) Color (6) |
| 54 | modelurcorr | Rest Frame SED Modeled ($u-r$) Color Corrected for Internal Extinction (7) |
| 55 | r50 | Half-light Radius in $r$ Band |
| 56 | r90 | 90% light Radius in $r$ Band |
| 57 | axialratio | Axial Ratio |
| 58 | mudelta | Morphological metric $\mu_{\Delta}$ (8) |
| 59 | mhidet | Galaxy H I Mass from 21 cm Detection (9) |
| 60 | e_mhidet | Statistical Uncertainty on Galaxy H I Mass from 21 cm Detection (Even if Confused) |
| 61 | mhilim | Upper Limit on H I Mass (10) |
| 62 | logmgas | Best Estimate of Atomic Gas Mass (11) |
| 63 | logmgastype | Type Flag for Atomic Gas Mass Estimate (12) |
| 64 | g3grp | G3 Group Identifier |

**Table 6**
(Continued)

| Column | Designation | Description |
|---|---|---|
| 65 | g3grpradeg | G3 Group Center R.A. J2000 |
| 66 | g3grpdedeg | G3 Group Center Decl. J200 |
| 67 | g3grpcz | G3 Group Center *cz* (13) |
| 68 | g3grpngi | Number of Giants in G3 Group |
| 69 | g3grpndw | Number of Dwarfs in G3 Group |
| 70 | g3grpabsrmag | G3 Group-integrated *r*-Band Absolute Magnitude |
| 71 | g3logmhvir | G3 HAM Log Group Mass (337b) (14) |
| 72 | g3logmh200 | G3 HAM Log Group Mass (200b) (15) |
| 73 | g3grpmhi | G3 Log Group-integrated H I Mass |
| 74 | g3fc | G3 Central Galaxy Flag (16) |
| 75 | g3skycutoffflag | RESOLVE-A Sky Cutoff Flag for G3 Groups (17) |
| 76 | involg3 | In-volume Flag for G3 Groups (18) |
| 77 | fofgrp | FoF Group Identifier |
| 78 | fofgrpradeg | FoF Group Center R.A. J2000 |
| 79 | fofgrpdedeg | FoF Group Center Decl. J2000 |
| 80 | fofgrpcz | FoF Group Center *cz* (13) |
| 81 | fofgrpn | Number of Galaxies in FoF Group |
| 82 | fofgrpabsrmag | FoF Group-Integrated *r*-Band Absolute Magnitude |
| 83 | foflogmhvir | FoF HAM Log Group Mass (337b) (14) |
| 84 | foflogmh200 | FoF HAM Log Group Mass (200b) (15) |
| 85 | fofgrpmhi | FoF Log Group-Integrated H I Mass |
| 86 | foffc | FoF Central Galaxy Flag (16) |
| 87 | fofskycutoffflag | RESOLVE-A Sky Cutoff Flag for FoF Groups (17) |
| 88 | involfof | In-volume Flag for FoF Groups (18) |

**Notes.** Only a portion of this table is shown here to demonstrate its form and content. The full table is available online in machine-readable format. (1) Value unchanged from Eckert et al. (2015). (2) Custom remeasured value from E15, provided without foreground extinction correction. (3) NUV foreground extinction corrections supersede values from E15, which were inadvertently tabulated (but not used) at 10 times the true value. (4) Log galaxy stellar mass derived using cz to calculate cosmological luminosity distances. (5) Absolute magnitude based on custom *r*-band photometry of E15, foreground extinction correction included, derived using cz to calculate cosmological luminosity distances. (6) Values include foreground extinction corrections and k-corrections, but not internal extinction corrections. (7) Values include foreground extinction corrections, k-corrections, and internal extinction corrections. (8) Morphological metric $\mu_{\Delta}$ defined in Kannappan et al. (2013), derived using cz to calculate cosmological angular diameter distances. (9) H I mass derived using cz to calculate cosmological luminosity distances; set to 0 if not detected or not observed. If the observation is confused, this column reports the total H I mass. (10) Upper limit on H I mass derived using cz to calculate cosmological luminosity distances; set to 0 if detected in H I or not observed. (11) Optimal atomic gas mass (1.4 × H I mass to include He) estimate based on detection, upper limit, confusion, and/or photometric gas fraction information; see Appendix A.6.2. (12) See Appendix A.6.2 for explanations of possible values. (13) Local Group-corrected recession velocity of group center, derived using cz. (14) HAM performed using the Tinker et al. (2008) halo mass function. (15) HAM performed using the Warren et al. (2006) halo mass function. (16) Set to 1 if galaxy is the group's central (most luminous in *r* band), or 0 otherwise. (17) RESOLVE-A sky cutoff flag, set to 1 for groups that have members both within RESOLVE-A and in the larger ECO survey outside RESOLVE-A (i.e., groups at the edges of the RESOLVE-A footprint). (18) Group in-volume flag; set to 1 if group satisfies $4500 < cz_{\rm grp}[{\rm km\ s^{-1}}] < 7000$. We do not tabulate invole17 for RESOLVE, as all quantities necessary to reconstruct prior RESOLVE membership have been published previously.

(This table is available in its entirety in machine-readable form.)

group velocity dispersion. As these association boundaries are typically larger than the FoF linking lengths, subfloor dwarf galaxies in E17 occupied statistically larger group-relative radii and velocities than dwarfs above the floor (which were included in FoF). To address this inconsistency and avoid the necessity of halo mass information, we have used an updated association algorithm for ECO DR3 and RESOLVE DR4. The new algorithm associates any subfloor dwarf to a group if it resides within the linking lengths of a galaxy in the original FoF group. In cases where a subfloor dwarf is linked to more than one FoF group, the algorithm selects the group corresponding to the previously FoF-grouped galaxy for which

$$\sqrt{\left(\frac{D_{\perp}}{b_{\perp}s}\right)^2+\left(\frac{D_{||}}{b_{||}s}\right)^2} \tag{A2}$$

is minimized, where $D_{\perp}$ and $D_{||}$ are the transverse and line-of-sight comoving distances, respectively, between the FoF-grouped galaxy and subfloor dwarf. We tested this technique by comparing subfloor dwarf and above-floor dwarf distributions of $R_{\rm proj,gal}$ and $\Delta v_{\rm proj,gal}$. Compared to E17, our updated method yields better agreement between the $R_{\rm proj,gal}$ and $\Delta v_{\rm proj,gal}$ distributions of above-floor and subfloor dwarfs, indicating our association method is more self-consistent with FoF group finding.

For both the G3 and FoF group catalogs, Table 5 provides for each ECO galaxy a group identifier, the group center coordinates, the number of group members, the group halo mass, and the group-integrated H I mass; a machine-readable table is available in the online version. It also tabulates the boundary completeness correction factor ccbremapped discussed in Section 2.4, which we have remapped to G3 groups using the ccb values and central galaxy names from the E17 FoF group catalog. All analogous group data for RESOLVE DR4 are located in Table 6, which also includes g3skycutoffflag and fofskycutoffflag.

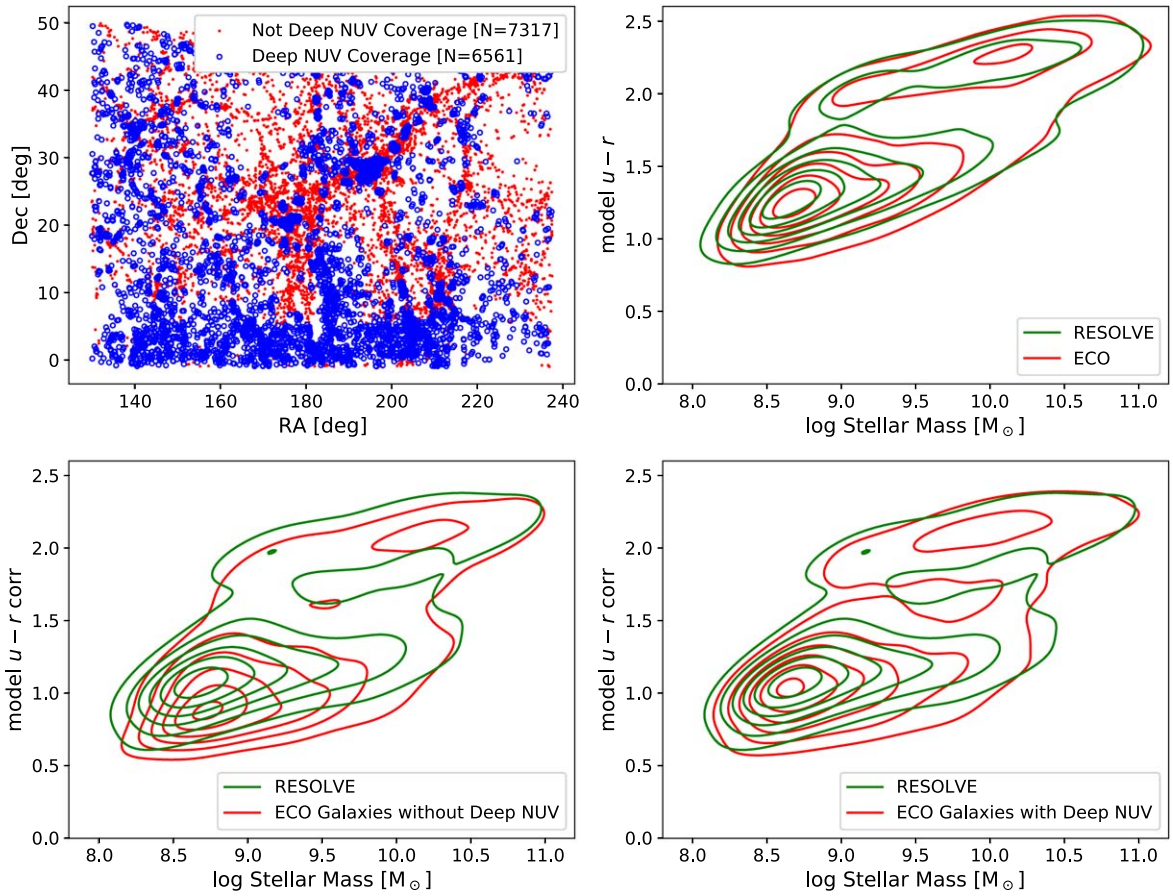


**Figure 19.** Inconsistency of extinction corrections performed with and without GALEX NUV data. Top left: GALEX NUV coverage of ECO, with "deep" defined as >1000 s exposure time. We note that RESOLVE (decl. 0°–5°) has full deep GALEX NUV coverage. Top right: model $u - r$ color vs. stellar mass for RESOLVE and ECO galaxies (all $M_r < -17.33$). There is no model $u - r$ offset between ECO and RESOLVE. Bottom left: model $u - r$ color corrected for internal extinction vs. stellar mass for RESOLVE (in which all galaxies have deep GALEX NUV) and for ECO (just galaxies without deep GALEX NUV). RESOLVE and ECO show a visible offset. Bottom right: model $u - r$ color corrected for internal extinction vs. stellar mass for RESOLVE and for ECO (just galaxies with deep GALEX NUV). The offset seen in the lower left panel disappears.

These columns flag galaxies in RESOLVE-A whose groups have members both within RESOLVE-A and in the larger ECO survey outside RESOLVE-A (i.e., groups at the edges of the RESOLVE-A footprint).

For both RESOLVE and ECO, we report group metrics for subfloor dwarfs as follows. If the subfloor galaxy is associated with a group, we directly inherit the properties (e.g., group $cz$ or halo mass) of its host group, where the calculation of these properties is limited strictly to the group members above the floor. However, if a subfloor dwarf is not associated with a group, we assume it to be isolated and report group metrics from galaxy properties (e.g., group $cz$ from its galaxy redshift). Following E16, we assign halo masses to subfloor objects by extrapolation of the ECO $M_{\rm halo} - L_{r,\rm tot}$ relation.

### A.6. H I Census

ECO DR3 includes a major addition of archival 21 cm data as well as value-added data such as upper limits and confusion flags (Appendix A.6.1). Beyond updates to the naming and organization conventions of H I parameters to match ECO DR3, there are no updates to the underlying RESOLVE atomic gas census as part of RESOLVE DR4. For RESOLVE DR4, our H I data columns are derived directly from Table 1 of Stark et al. (2016), hereafter S16. Specifically, S16's `F21` (total 21 cm flux) is now provided in terms of H I mass via the relation $M_{\rm H\,I} = [2.36 \times 10^5/(1+z)^2]D_L^2 F_{21}$, where $D_L$ is the cosmological luminosity distance to the galaxy in Mpc calculated from `cz`, and $F_{21}$ is the total 21 cm observed frame flux in units of Jy km s$^{-1}$ (Meyer et al. 2017). To improve clarity and prevent misuse of our upper limits, we provide these H I masses for detections and upper limits separately, in columns `mhidet` and `mhilim`. For consistency, S16's `e_F21` (uncertainty on `F21`) is also provided in RESOLVE DR4 in terms of H I mass uncertainty; uncertainties on `mhidet` are reported in the column `e_mhidet`. We do not tabulate other H I data columns from S16 because we did not alter their values or organization. We refer our reader to S16 or to the online RESOLVE database (https://resolve.astro.unc.edu/pages/database.php) for additional data regarding the RESOLVE H I observations, such as confusion information.

For both ECO DR3 and RESOLVE DR4, we add a column `logmgas` based a new algorithm for estimating a single "best" estimate of a galaxy's atomic gas mass, as described in Appendix A.6.2 and used in our Section 5 analysis. All H I parameters for ECO DR3 are provided in Table 7, and the new and reorganized H I parameters for RESOLVE DR4 are provided in Table 6.

#### A.6.1. ECO—ALFALFA 100 Cross-match

The ECO DR2 H I catalog was constructed primarily from a cross-match with the ALFALFA 40% release

**Table 7**
ECO DR3 21 cm Catalog

| Column | Designation | Description |
|---|---|---|
| 1 | name | ECO Galaxy Identifier |
| 2 | mhidet | Galaxy H I Mass from ALFALFA-100 or RESOLVE Detection (1) |
| 3 | e_mhidet | Statistical Uncertainty on Galaxy H I Mass from 21 cm Detection (Even if Confused) |
| 4 | confused | Confusion Flag (2) |
| 5 | mhilim | Upper Limit on H I Mass from ALFALFA-100 or RESOLVE (3) |
| 6 | limsigma | Integrated rms Noise for Nondetection (4) |
| 7 | limmult | Upper Limit Level, 3 or 5 (5) |
| 8 | hirms | rms Noise Level of H I Spectrum (6) |
| 9 | hitelescope | Source of H I Data (7) |
| 10 | peaksnhi | Peak S/N of 21 cm Line Detection |
| 11 | logmgas | Best Estimate of Atomic Gas Mass (8) |
| 12 | logmgastype | Type Flag for Atomic Gas Mass Estimate (9) |
| 13 | mhideta100 | Galaxy H I Mass from ALFALFA-100 21 cm Detection (1) |
| 14 | e_mhideta100 | Statistical Uncertainty on ALFALFA-100 H I Mass (Even if Confused) |
| 15 | confuseda100 | Confusion Flag for ALFALFA-100 Detection (2) |
| 16 | agcnra100 | Arecibo General Catalog Number |
| 17 | mhilima100 | Upper Limit on ALFALFA-100 H I Mass (3) |
| 18 | hirmsa100 | rms Noise Level of H I Spectrum (6) |
| 19 | peaksnhia100 | Peak S/N of ALFALFA-100 21 cm Line Detection |
| 20 | logmgasa100 | Best Estimate of Atomic Gas Mass from ALFALFA-100 Data (8) |
| 21 | logmgastypea100 | Type Flag for Atomic Gas Mass Estimate from ALFALFA-100 Data (9) |

**Notes.** Only a portion of this table is shown here to demonstrate its form and content. The full table is available online in machine-readable format. As described in Appendix A.6.1, ECO DR3 H I data are provided in two sets of H I-related parameters: parameters combining ALFALFA-100 and deeper RESOLVE observations (columns 2–12), and parameters based exclusively on ALFALFA-100 data (columns 13–21). (1) H I mass derived using cz to calculate cosmological luminosity distances; set to 0 if not detected or not observed. If the observation is confused, this column reports the total H I mass. (2) 1 = observation likely confused; 0 = not confused or not observed. (3) Upper limit on H I mass derived using cz to calculate cosmological luminosity distances; set to 0 if detected in H I or not observed. (4) Based on estimated rms noise level and estimated $W_{20}$ assuming 10 km s$^{-1}$ channels; set to 0 for detections. Upper limits computed at three or five times this value. (5) Level of upper limit: set to 3 for RESOLVE sources and 5 for rest of ECO in ALFALFA-100 field; set to 0 for detections. (6) Measured from 21 cm spectra for RESOLVE-A galaxies, drawn from ALFALFA-100 catalog for detected ECO galaxies, or estimated from Figure 21(b) for nondetections. (7) Possible values: GBT = Green Bank Telescope (S16), AO = Arecibo Observatory (S16), ALFALFA=ALFALFA H I Survey (Haynes et al. 2011, 2018), and S05 = Springob et al. (2005) catalog. (8) Optimal atomic gas mass (1.4 × H I mass to include He) estimate based on detection, upper limit, confusion, and/or photometric gas fraction information; see Appendix A.6.2. (9) See Appendix A.6.2 for explanations of possible values.

(This table is available in its entirety in machine-readable form.)

(Haynes et al. 2011), with an additional ∼350 observations obtained via overlap with the RESOLVE atomic gas census (Stark et al. 2016; hereafter S16). In ECO DR3, we update the H I data set using the full ALFALFA 100% release ($\alpha$100; Haynes et al. 2018) and add upper limits and confusion flags. Figure 20 shows the on-sky map of ECO and illustrates the substantial amount of new H I detections or upper limits introduced in ECO DR3.

To match 21 cm sources from the $\alpha$100 catalog, we perform an on-sky and line-of-sight cross-match following the methodology of S16. In the plane of the sky, we use a matching radius of 4′, which roughly matches the ALFA half-power beamwidth ($3\farcm3 \times 3\farcm8$). In the line of sight, we consider optical and radio sources to be matched when the optical redshift window, $cz \pm \sigma_{cz} \pm (\Delta cz)/2$, overlaps with the 21 cm redshift window, $cz_{\rm H\,I} \pm W_{50}/2$, where $W_{50}$ is the 21 cm linewidth at 50% peak level, and $\Delta cz$ is the galaxy assumed linewidth (see description below). We simultaneously generate value-added confusion flags during the matching. We mark galaxies as confused if there are multiple optical sources within the spatial matching radius of a single ALFALFA H I source and with redshift matching windows that overlap with the 21 cm profile of the same single ALFALFA source. Unlike for RESOLVE, no deconfusion was attempted for ECO.

The assumed linewidth parameter $\Delta cz$ is meant to extend the optical matching window such that gas can be detected at higher or lower velocities. Therefore, the assumed linewidth should reflect the largest true linewidth that we would reasonably expect most galaxies to have, so that the matching algorithm is inclined toward flagging more targets as confused rather than missing cross-matches altogether. Figure 21(a) shows the distribution of $W_{50}$ values from $\alpha$100 in the redshift range of ECO. The plot illustrates that 90% of ALFALFA-detected sources have $W_{50}$ less than 300 km s$^{-1}$, so we opt to use this value for $\Delta cz$. E16 previously used $\Delta cz = 200$ km s$^{-1}$ for ECO DR2, though the full $\alpha$100 data set suggests 200 km s$^{-1}$ might miss cross-matches.

Following the cross-match and generation of confusion flags, we compute upper limits on the H I masses of ECO galaxies not detected by ALFALFA in its survey footprint, $112\fdg5 < {\rm RA} < 247\fdg5$ and $0° < {\rm decl.} < 36°$. We compute $5\sigma$ upper limit masses using Equation (1) of S16. Upper limits are based on rms noise estimates from the decl.-dependent model shown in Figure 21(b). We estimate the number of channels enclosed in the line profile by assuming a velocity resolution of 10 km s$^{-1}$ and the estimate of the individual galaxy's $W_{20}$, the 21 cm linewidth at 20% peak level. The $W_{20}$ values are computed as $W_{20} = W_{50} + 20$ km s$^{-1}$ (Haynes et al. 1999), with

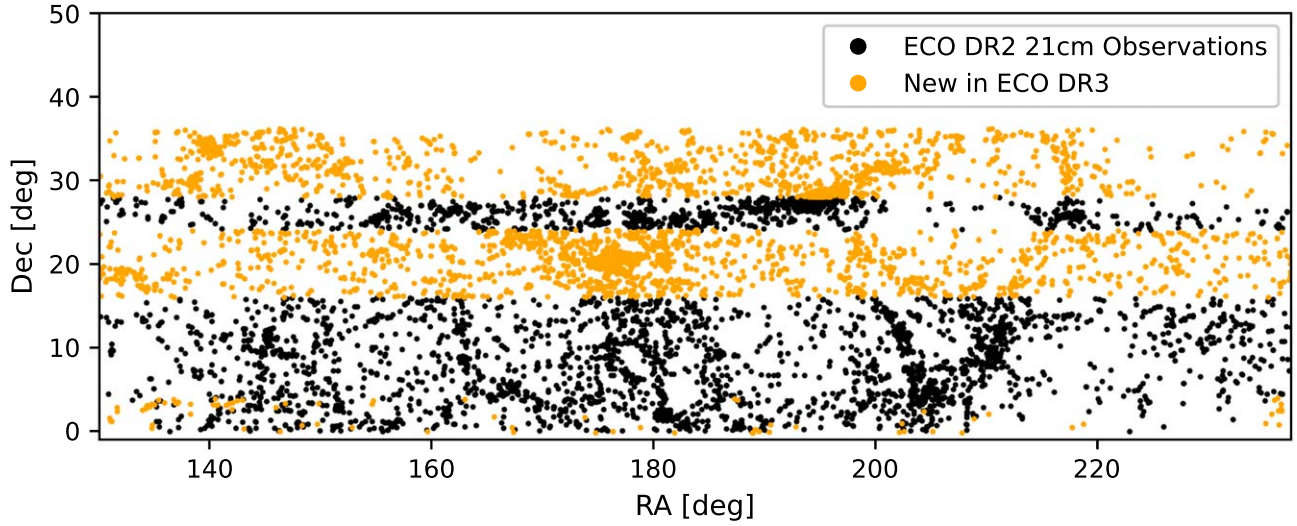


**Figure 20.** Sky map of the ECO survey illustrating new H I data in ECO DR3 (orange) compared with ECO DR2 (black; Eckert et al. 2017). Orange points within the RESOLVE-A footprint (<5° decl.) represent new H I data for ECO galaxies located at lower redshifts than RESOLVE-A (which is limited to 4500–7000 km s$^{-1}$, while ECO reaches down to 3000 km s$^{-1}$).

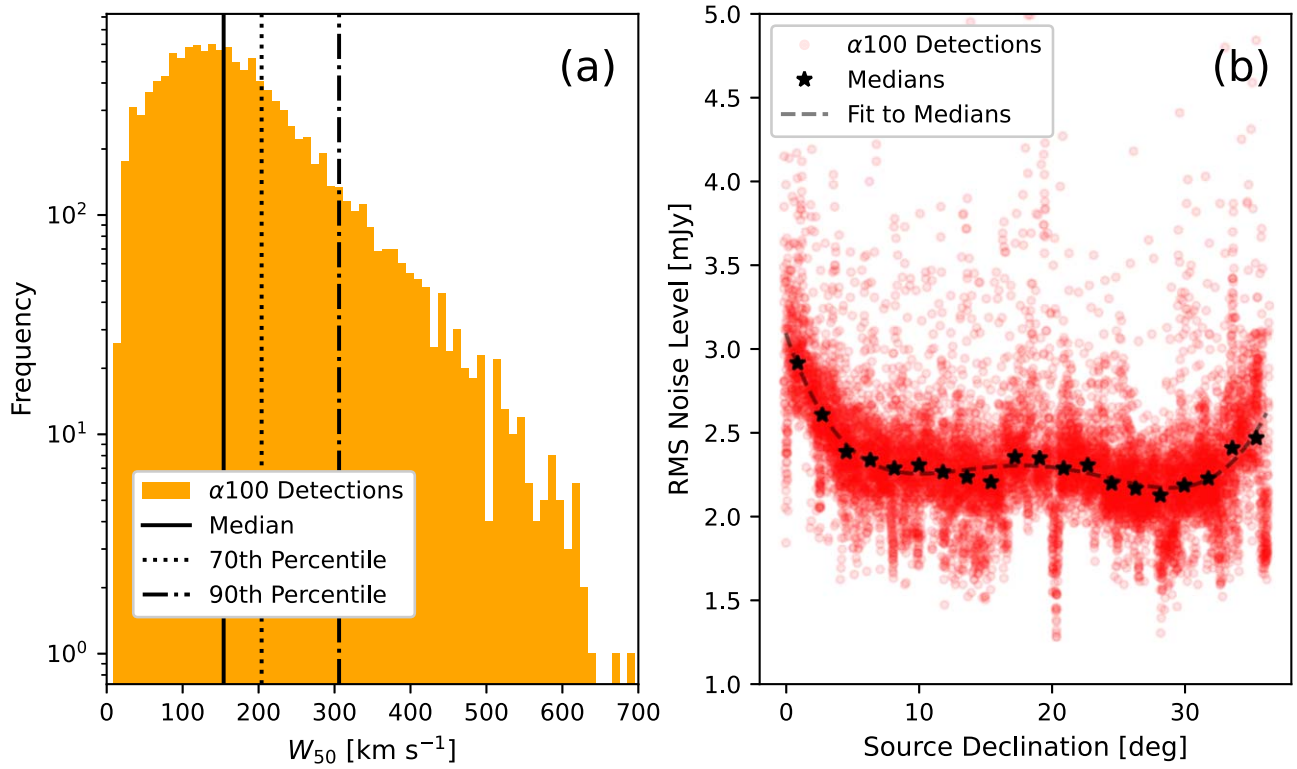


**Figure 21.** Linewidths and rms noise levels for ALFALFA-100 detections (Haynes et al. 2018) in the $2530 < v_{\rm H\,I}$[km s$^{-1}$] $< 7470$ range, to be used in optical–H I cross-matching and in calculating value-added H I upper limits. (a) Distribution of 21 cm 50% peak linewidths ($W_{50}$) of galaxies detected in ALFALFA-100, with black lines marking the 50th, 70th, and 90th percentiles. (b) decl.-dependent rms noise from 21 cm spectra of galaxies detected in ALFALFA-100. Black stars are binned medians and the dashed line is a quartic least-squares fit to those medians.

$W_{50}$ estimated from the baryonic Tully–Fisher relation of Kannappan et al. (2013).

From the cross-match, value-added confusion flags, and value-added upper limits, we compile the new H I catalog for ECO DR3. Unlike $\alpha$40, $\alpha$100 covers the RESOLVE-A overlap of ECO, where there are also deeper RESOLVE observations from Arecibo and Green Bank (S16). Thus, for some of these galaxies, there are two possible sources of H I data. To provide data from both sources, the ECO DR3 database includes two sets of H I-related parameters. The primary H I parameters (e.g., `mhidet`) mix ALFALFA-100 and RESOLVE observations, providing the highest-quality H I data available for each galaxy. However, these parameters reflect spatially inhomogeneous depth and differing upper limit thresholds over the ECO sky—RESOLVE upper limits are computed at $3\sigma$ (not $5\sigma$) because H I observations were carried out at known optical galaxy positions. In contrast, the $\alpha$100-based parameters (e.g., `mhidet_a100`) exclude deeper RESOLVE observations and

provide uniform-depth ALFALFA data over the ECO sky, with all value-added upper limits computed at the $5\sigma$ level.

#### A.6.2. *Best* $\log M_{\rm H\,I}$ *Algorithm*

For both RESOLVE DR4 and ECO DR3, we compute a "best" $\log M_{\rm H\,I}$ estimate `logmgas` and flag `logmgastype` for each galaxy, provided in Tables 6 and 7, by combining detection, upper limit, confusion, and photometric gas fraction information. We adopt as-is all strong detections and strong upper limits ($1.4M_{\rm H\,I}/M_* < 0.05$), treating both $3\sigma$ and $5\sigma$ strong upper limits as if they were detections at the low level of the limit. Additionally, we adopt as-is successfully deconfused detections, as well as confused H I detections for which $1.4M_{\rm H\,I}/M_* < 0.1$, even if deconfusion was unsuccessful or not attempted. We regard deconfusion as "successful" if `eF21corr-sys / F21 > 0.25` and `F21corr <=F21` (see Table 1 of S16 for these data). All such as-is adoptions are assigned `logmgastype=0`. For missing H I observations (`logmgastype=3`), weaker upper limits (`logmgastype=2`), or in other confused cases where deconfusion was unsuccessful or not attempted (`logmgastype=1`), we estimate the galaxy H I mass using the photometric gas fraction technique (Kannappan 2004, as updated in E15), constraining these estimates using our weak upper limits or confused detections when possible. We obtain a probability distribution for possible H I gas-to-stellar mass ratios from the public code of E15 using the galaxy's $u-J$ color and axial ratio, multiply by the galaxy's stellar mass, then exclude the portion of the distribution above either the observed weak upper limit H I mass or the total mass of a confused detection, as relevant. The median of the leftover distribution is chosen as the best estimate of the galaxy's H I mass. The algorithm described here is publicly available as part of the RESOLVE Database Tutorial (https://github.com/resolvesurvey/database-tutorial).

## ORCID iDs

Zackary L. Hutchens https://orcid.org/0000-0002-8574-5495
Sheila J. Kannappan https://orcid.org/0000-0002-3378-6551
Kathleen D. Eckert https://orcid.org/0000-0002-1407-4700
Andrew J. Baker https://orcid.org/0000-0002-7892-396X
Kelley M. Hess https://orcid.org/0000-0001-9662-9089
Mark A. Norris https://orcid.org/0000-0002-7001-805X
Darren Croton https://orcid.org/0000-0002-5009-512X

## References

Abadi, M. G., Moore, B., & Bower, R. G. 1999, MNRAS, 308, 947
Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634
Ai, M., & Zhu, M. 2018, ApJ, 862, 48
Baugh, C., Gonzalez-Perez, V., Lagos, C. D., et al. 2019, MNRAS, 483, 4922
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2012, ApJ, 762, 109
Bekki, K., Couch, W. J., & Shioya, Y. 2002, ApJ, 577, 651
Bentley, J. L. 1975, Commun. ACM, 18, 509
Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006, ApJS, 167, 1
Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349
Blanton, M. R., & Berlind, A. A. 2007, ApJ, 664, 791
Blyth, S. L., Baker, A. J., Holwerda, B. W., et al. 2016, in Proc. MeerKAT Science: On the Pathway to the SKA, 4
Calette, A. R., Rodríguez-Puebla, A., Avila-Reese, V., & Lagos, C. d. P. 2021, MNRAS, 506, 1507
Castorina, E., & Villaescusa-Navarro, F. 2017, MNRAS, 471, 1788
Chauhan, G., Lagos, C. d. P., Stevens, A. R., et al. 2020, MNRAS, 498, 44
Chauhan, G., Lagos, C. d. P., Stevens, A. R., et al. 2021, MNRAS, 506, 4893
Chen, J., Zhang, P., Zheng, Y., Yu, Y., & Jing, Y. 2018, ApJ, 861, 58
Dalcanton, J. J., Yoachim, P., & Bernstein, R. A. 2004, ApJ, 608, 189
Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Natur, 433, 604
Duarte, M., & Mamon, G. A. 2014, MNRAS, 440, 1763
Eckert, K. D., Kannappan, S. J., Lagos, C. d. P., et al. 2017, ApJ, 849, 20
Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2015, ApJ, 810, 166
Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2016, ApJ, 824, 124
Garnett, D. R. 2002, ApJ, 581, 1019
Gerke, B. F., Newman, J. A., Faber, S., et al. 2007, MNRAS, 376, 1425
Guo, H., Jones, M. G., Haynes, M. P., & Fu, J. 2020, ApJ, 894, 92
Guo, H., Zheng, Z., Zehavi, I., et al. 2015, MNRAS, 453, 4368
Haynes, M. P., Giovanelli, R., Chamaraux, P., et al. 1999, AJ, 117, 2039
Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, ApJ, 861, 49
Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, AJ, 142, 170
Hess, K. M., & Wilcots, E. M. 2013, AJ, 146, 124
Hood, C. E., Kannappan, S. J., Stark, D. V., et al. 2018, ApJ, 857, 144
Huchra, J., & Geller, M. 1982, ApJ, 257, 423
Kannappan, S. J. 2004, ApJL, 611, L89
Kannappan, S. J., Guie, J. M., & Baker, A. J. 2009, AJ, 138, 579
Kannappan, S. J., Stark, D. V., Eckert, K. D., et al. 2013, ApJ, 777, 42
Kannappan, S. J., & Wei, L. H. 2008, in AIP Conf. Proc. 1035, THE EVOLUTION OF GALAXIES THROUGH THE NEUTRAL HYDROGEN WINDOW (Melville, NY: AIP), 163
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 54
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Kim, H.-S., Wyithe, J. S. B., Baugh, C. M., et al. 2017, MNRAS, 465, 111
Kourkchi, E., & Tully, R. B. 2017, ApJ, 843, 16
Li, X., Li, C., Mo, H., Xiao, T., & Wang, J. 2022, ApJ, 941, 48
Lim, S., Mo, H., Lu, Y., Wang, H., & Yang, X. 2017, MNRAS, 470, 2982
Malavasi, N., Aghanim, N., Douspis, M., Tanimura, H., & Bonjean, V. 2020, A&A, 642, A19
Meyer, M., Robotham, A., Obreschkow, D., et al. 2017, PASA, 34, e052
Moffett, A. J., Kannappan, S. J., Berlind, A. A., et al. 2015, ApJ, 812, 89
Moré, J. J. 1978, Numerical analysis (Berlin: Springer), 105
Obuljen, A., Alonso, D., Villaescusa-Navarro, F., Yoon, I., & Jones, M. 2019, MNRAS, 486, 5124
Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016, A&A, 594, A1
Rodriguez, F., & Merchán, M. 2020, A&A, 636, A61
Roychowdhury, S., Meyer, M. J., Rhee, J., et al. 2022, ApJ, 927, 20
Skibba, R. A., van den Bosch, F. C., Yang, X., et al. 2011, MNRAS, 410, 417
Spina, B., Porciani, C., & Schimd, C. 2021, MNRAS, 505, 3492
Springel, V. 2005, MNRAS, 364, 1105
Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, ApJS, 160, 149
Stark, D. V., Kannappan, S. J., Eckert, K. D., et al. 2016, ApJ, 832, 126
Stierwalt, S., Liss, S. E., Johnson, K. E., et al. 2017, NatAs, 1, 0025
Stothert, L., Norberg, P., & Baugh, C. 2019, MNRAS Lett., 485, L126
Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810
Taylor, M. 2017, Informatics, 4, 18
Tempel, E., Kipper, R., Tamm, A., et al. 2016, A&A, 588, A14
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Tinker, J. L. 2007, MNRAS, 374, 477
Tully, R. B., Rizzi, L., Dolphin, A., et al. 2006, AJ, 132, 729
van den Bosch, F. C., Weinmann, S. M., Yang, X., et al. 2005, MNRAS, 361, 1203
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Warren, M. S., Abazajian, K., Holz, D. E., & Teodoro, L. 2006, ApJ, 646, 881
Yang, X., Mo, H., Jing, Y., & Van Den Bosch, F. C. 2005, MNRAS, 358, 217
York, D. G., Adelman, J., Anderson, J. E., Jr, et al. 2000, AJ, 120, 1579
Zabludoff, A. I., & Mulchaey, J. S. 1998, ApJ, 496, 39
Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760