Title
Modeling the Formation and Evolution of Satellite Galaxies in Cosmological Simulations

Author
Santistevan, Isaiah

Publication Date
2023

Peer reviewed|Thesis/dissertation

Modeling the Formation and Evolution of Satellite Galaxies in Cosmological Simulations

By

ISAIAH BENJAMIN SANTISTEVAN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Andrew Wetzel, Chair

Tucker Jones

Christopher Fassnacht

Committee in Charge

2023

To all of my (great) grandparents, past and present,
for persevering throughout the tough New Mexican history.

Contents

Abstract vi

Acknowledgments vii

Chapter 1. Introduction 1

1.1. Hierarchical Structure Formation 1

1.2. Satellite Galaxies 5

1.3. Modeling Satellite Galaxy Orbits 8

1.4. Cosmological Simulations 9

1.5. Dissertation Outline 12

Chapter 2. The Formation Times and Building Blocks of Milky Way-mass Galaxies in the FIRE Simulations 14

2.1. Abstract 14

2.2. Introduction 15

2.3. Methods 20

2.4. Results 27

2.5. Summary and Discussion 48

Acknowledgements 56

Chapter 3. The origin of metal-poor stars on prograde disk orbits in FIRE simulations of Milky Way-mass galaxies 58

3.1. Abstract 58

3.2. Introduction 59

3.3. Methods 63
3.4. Results 70
3.5. Summary & Discussion 96
Acknowledgements 106

Chapter 4. Orbital dynamics and histories of satellite galaxies around Milky Way-mass galaxies in the FIRE simulations 108
4.1. Abstract 108
4.2. Introduction 109
4.3. Methods 112
4.4. Results 121
4.5. Summary & Discussion 148
Acknowledgements 157

Chapter 5. Modeling orbital histories of satellites of Milky Way-mass galaxies: testing static host potential models against cosmological simulations 158
5.1. Abstract 158
5.2. Introduction 159
5.3. Methods 164
5.4. Results 173
5.5. Summary & Discussion 209
Acknowledgements 217

Chapter 6. Conclusion 218
6.1. Previous Work 218
6.2. Future Work 220

Appendix A. Chapter 4 Appendices 223
A.1. Trends with Peak Halo Mass 223

A.2. Baryonic versus dark-matter-only simulations 227

Appendix B. Chapter 5 Appendix 231

B.1. Modeling the Mass/Potential Profiles 231

Bibliography 237

Abstract

Studying various properties of stellar populations within a galaxy provides rich insight into its formation, and current surveys enable us to do so with the goal of reconstructing their formation histories. However, it is ambiguous as to when the main progenitor of a galaxy like the Milky Way (MW) formed, and what its constituent mergers were like. In the first half of this work, I use the FIRE-2 cosmological simulations to study these low-mass galaxies at early times, and their later mergers. I quantify the formation times of MW/M31-mass galaxies based on different definitions of formation, as well as the mass function of its constituent building blocks. I find that the MW-mass galaxies in Local Group (LG)-like pairs form earlier, highlighting the role that environment may play in galaxy formation. I also find a similar ubiquitous feature of metal-poor stars in the disks of the MW-mass hosts on preferentially prograde orbits, which are also found in the MW. These stars largely originate ex-situ, from a gas-rich merger of mass similar to the Magellanic Clouds, that sometimes shapes the orientation of the stellar disk.

We call the gravitationally bound low-mass galaxies that survive the hierarchical formation process satellite galaxies. The satellite galaxies in the LG are the *only* low-mass galaxies that we can extensively study through fully resolved stellar populations and their full phase-space coordinates. In the second half of this work, I study the infall and orbit histories of satellite galaxies around the MW-mass hosts in the simulations. I examine trends in their present-day orbital dynamics, including total velocity, specific angular momentum, and total energy, as well as their orbital histories versus present-day distance, stellar mass, and the lookback time of infall. Surprisingly, the majority of satellites that experience multiple orbits have larger recent pericenters than their minimum, contradictory to the assumption that satellite orbits always shrink. Finally, I compare the satellite galaxy orbits in a static, axisymmetric MW-mass potential to the simulations to quantify the extent to which this common orbit modeling technique works. I find that orbital energy and specific angular momentum are not always conserved, and orbit properties that occurred recently, such as pericenters, are better recovered than events further in the past, such as infall time.

Acknowledgments

This twelve year journey, from community college where I re-learned basic high school math to graduate school now, is an endeavor I could not have achieved on my own. I owe my success to the collective encouragement, support, and mentoring from just about everyone in my life.

I would literally and figuratively not be here if not for both of my parents, Cynthia Sanchez and Ben Santistevan. Thank you both for your endless well of love and support, and for both embodying what it means to be kind, curious, and resilient. If I were to live a thousand lifetimes, it still would not be long enough to repay all that you have done for me, nor long enough for me to express my gratitude. To my sister, Jennifer, and her family, thank you for always being one of my biggest supporters and confidants when life gets tough.

To all of my grandparents and great-grandparents, Clara Sanchez, Ernest and Gloria Santistevan, Ben and Lucy Maestas, Honey and Popo Garcia, and Armando Santistevan. Thank you for showing me what it means to endure through the culturally and environmentally tough New Mexican history, and for the love, lessons, and stories you shared with me.

To my advisor and mentor, Andrew Wetzel, I could not have done any of this without your guidance, patience, and belief that I can succeed. I feel honored to have studied under your watch, and I find myself fortunate to have absorbed even a fraction of your wisdom. I looked forward to all of our one-on-one meetings to hear your insight on all things scientific and everyday. Thank you for teaching me how to write, through code and words (despite my occasional passive voice usage), and for what it means to be rigorous. I am further blessed to have been welcomed by your family and to have spent so much time with Theodore.

To my love, Lilli. Thank you for your unwavering love and support and for being a beacon in even the toughest of times. I am truly blessed for the life and memories we have built together and for the many more laughs we will share in the future. My life is far more rich with you in it.

To all of the members in my research group, past and present. Jenna, Matt, Pratik, Fiona, Preet, Megan, Sarah, and Samantha, thank you all for your ear and mind when I needed help and honest

criticism when I needed it most. Thank you for also sharing your humor and struggles which made grad school more personable and memorable.

To my advisors and mentors, present and past. Tucker Jones and Chris Fassnacht, thank you for your wisdom and guidance throughout grad school. Erik Tollerud and Robyn Sanderson, thank you for your support throughout the years and for always offering critiques that have only made the papers we wrote even better. Matt Richter, thank you for the opportunity to learn about your research and for showing me the lab at the NASA Ames Facility. Ylva Pihlström and Loránt Sjouwerman, thank you for sharing your humor and your enthusiasm for your work while I was still an undergraduate. Richard Rand, thank you for taking me on when I transferred from CNM, for giving me the opportunity to contribute to your research over a summer, and for your guidance and advice in transferring to grad school. Kurtis Williams, thank you for taking me in for my first astronomy research endeavor and for the thoughtful check-in when I submitted my second paper. Judy Lalani, thank you for all of the jokes during class, and for first instilling the mathematical rigor necessary to succeed in STEM.

To all of my friends back home in New Mexico. Carlos Apodaca, thank you for being the brother I never had, for the countless hours of your friendship, and for your unwavering support over the last 16 years. Justin Salazar, thank you for your incredible quick wit, corny jokes, and for your stalwart friendship. Jason Orona, thank you for the many laughs we shared together and for welcoming me into your home when I visit. Bronson Roybal, thank you for all of the interesting conversations we shared. Estevan Robles, thank you for your positivity and for the fantastic art you share with the world. Christian Roberts, thank you for the many days and nights we spent in the physics building doing homework or playing ping pong, and for the many adventures we shared throughout the Southwest. Cameron Trapp, Merrick Chalk, Michael Brown, and Alex Cordova, thank you for your companionship and humor as we made it through the UNM physics department mostly unscathed.

Finally, to all of the new friends and colleagues that I made here at UCD. Jeff Harvey, Mikhael Semaan, and Kendall Berry, thank you for your kindness and friendship and the many meals we shared together. Victoria, Jessie, Morgane, Geoff, Azalee, Imran, Brian, Keerthi, and Ryan, thank you all for making the astronomy department welcoming, inclusive, and for sharing a laugh or story in our offices.

CHAPTER 1

Introduction

One of the most fundamental areas of research in astronomy is understanding how galaxies form. From the initial dark matter overdensities at exceptionally early times to the present-day Universe as we know it, many physical processes occurred to produce a galaxy like our own, the Milky Way (MW). We see signs of this formation process within the stellar halo and disk of the MW, through stars that share similar kinematics, elemental abundances, and/or ages. We additionally observe coherent formation relics, such as stellar streams that are in the process of disrupting and satellite galaxies that continue to orbit around the MW. Because these satellite galaxies survived the formation history of the MW, understanding their dynamics today gives us insight into their co-formation and evolution with the MW. Although we only see the MW as it is today, at one snapshot in time, simulations that model the formation and evolution of galaxies through a vast array of complicated physics are advanced enough so that we may study galaxies similar to our own across cosmic time.

1.1. Hierarchical Structure Formation

Eggen et al. (1962) first suggested that the MW initially formed from the collapse of a giant monolithic cloud of gas that subsequently formed stars on highly radial orbits, and after this initial gas had settled into a disk, a second generation of stars then formed on disk-like, circular orbits. However, other studies, such as Ostriker & Tremaine (1975); Searle & Zinn (1978), suggested that stars in the stellar halo of the MW formed in other low-mass galaxies that subsequently accreted into the MW after the main galaxy started forming, which led to the more commonly accepted theory of galaxy formation: that galaxies formed 'hierarchically', through many mergers of lower mass galaxies. Following initial studies suggesting hierarchical galaxy formation, (for example Gunn &

Gott, 1972; Press & Schechter, 1974), other studies suggested dark matter halos form in a similar manner (for example Fillmore & Goldreich, 1984), where initial dark matter over-densities in the Universe collapsed to form dark matter halos which act as sites of galaxy formation. It is then the dissipational collapse of gas within these dark matter halos that sparks initial galaxy formation (for example Rees & Ostriker, 1977; White & Rees, 1978; Fall & Efstathiou, 1980; Mo et al., 1998). However, the processes involved in galaxy formation are complex (Freeman & Bland-Hawthorn, 2002; Springel & Hernquist, 2005; Robertson et al., 2006).

1.1.1. Mergers. Mergers play both a destructive and constructive role in galaxy formation. In the destructive sense, larger ellipticals form from a merger between two massive galaxies of roughly equal mass, and any spiral structure present in either component is subsequently destroyed (Toomre & Toomre, 1972; Barnes & Hernquist, 1992). Bournaud et al. (2005) suggest that the outcome of a given merger between two galaxies depends on the initial mass ratio, with disk-like galaxies forming or growing from ratios of 4.5 : 1 or larger, and ellipticals forming from ratios closer to unity. Mergers sometimes result in the *formation* of a disk, or change the properties of existing disks (Bournaud et al., 2004; Springel & Hernquist, 2005; Robertson et al., 2006; Peschken et al., 2020).

Stars that orbit within the stellar halo of the MW are typically old and metal-poor. Many studies investigated the hierarchical formation of the stellar halo, and suggest that it formed through the disruption and accretion of low-mass galaxies at early times (Bullock et al., 2001; Helmi, 2008; Johnston et al., 2008; Cooper et al., 2010). Simulation based studies suggest that the halo formed primarily through the accretion of $1 - 2$ satellite galaxies with stellar masses $M_{\rm star} \gtrsim 10^8\,{\rm M}_\odot$ (Deason et al., 2016; Monachesi et al., 2019; Santistevan et al., 2020), and recent observational studies suggest similar results (Helmi et al., 2018; Belokurov et al., 2018; Naidu et al., 2020). Because the stars in the thick disk of the MW overlap in their kinematics and elemental abundances with stars in the stellar halo, some studies suggest similar formation mechanisms through low-mass galaxy accretion (Walker et al., 1996; Villalobos & Helmi, 2008). However, other studies suggest

that the stars in the thick disk were either kinematically heated or that the disk formed thick initially and settled over time (Loebman et al., 2011; Bird et al., 2013; Grand et al., 2016).

Finally, large amounts of rapid star formation may occur in the resulting merger if either of the galaxies contain a significant amount of gas, in events called starbursts (Barnes, 2004; Lin et al., 2008; Kim et al., 2009). Gas-rich mergers therefore increase the star formation rates, where as gas-poor mergers do not form starbursts (Athanassoula et al., 2016; Pearson et al., 2019). However, because galaxies were frequently merging in the early Universe, gas-rich mergers become less common than gas-poor mergers (Lin et al., 2008, 2010).

1.1.2. Imprints of mergers. A natural consequence of the hierarchical structure formation paradigm of galaxy formation is that not only are mergers between galaxies common during formation, but also that lower-mass mergers can sometimes leave imprints within the more massive host through the kinematics of their stars or stellar elemental abundances. Thus, we can sometimes discover the remnants of a merger within the larger galaxy if we happen to find a subset of stars with self-similar orbits and/or elemental abundances, even if the stars from the merger do not look like a distinct galaxy or star cluster. Recent advancements in technology have allowed astronomers to pinpoint which stars may have belonged together before merging into the MW. The Gaia mission has been revolutionary in providing positions, magnitudes, and velocities for over 1 billion stars within the MW, along with its satellite galaxies and star clusters (Gaia Collaboration et al., 2018). Using this wealth of information, Helmi et al. (2018) and Belokurov et al. (2018) discovered a population of stars in the solar neighborhood with similar radial and tangential velocities, and similar iron abundances, [Fe/H] ~ -1.6, which suggested that these stars shared a similar origin. This previous merger has since been named the 'Gaia-Enceladus' merger, but is also sometimes referred to as the 'Gaia-Sausage' or 'Gaia-Enchilada' merger based on how the distribution of stars looks like in velocity-space.

Other spectroscopic focused surveys, such as RAVE (Steinmetz et al., 2006), LAMOST (Li et al., 2015), APOGEE (Majewski et al., 2017), and GALAH (De Silva et al., 2015), provide

elemental abundances for many stars in the MW, which help with disentangling substructures of old and metal-poor stars within the halo (e.g. Yuan et al., 2019). Naidu et al. (2020) recently used data from the 'Hectochelle in the Halo at High Resolution' (H3) survey (Conroy et al., 2019) to identify four new substructures within the MW halo, each with unique iron abundances and orbital energies.

In another study, Sestito et al. (2019) compiled many of the most metal-poor stars, [Fe/H] < -2.5, from the *Pristine* (Starkenburg et al., 2017b) and LAMOST surveys, and discovered that for stars that orbit and currently reside close to the MW disk, the majority of them were orbiting in the prograde direction. This overabundance in prograde-to-retrograde stars persists down to [Fe/H] < -4 (Sestito et al., 2019), and is seen in simulations of MW-mass galaxies which suggest that these prograde, metal-poor stars originate from an gas-rich merger, with a stellar mass comparable to the Large or Small Magellanic Cloud (LMC/SMC), early in the formation of the MW (Sestito et al., 2021; Santistevan et al., 2021). A complete understanding of each component that contributed to the growth of the MW may not be possible in the near future, however, these new surveys now allow us to extract some mergers in the recent past.

1.1.3. Stellar streams. While some low-mass galaxies or star clusters completely tidally disrupt when merging into more massive galaxies, this is certainly not the end result of all mergers. The stellar streams in the outer halo of the MW provide a unique opportunity to study mergers before the stars completely phase mix with the other stars in the halo. One of first observations of such a system was of the low-mass satellite galaxy called 'Sagittarius', which orbits the MW with an estimated radial velocity of $\sim 140\,\mathrm{km\,s^{-1}}$ and distance of 24 kpc Ibata et al. (1994). Because of the spatial extent of the stars in Sagittarius, ~ 3 kpc, it was initially thought to be in the process of tidally disrupting, and later studies confirmed a stream component to the galaxy and both were orbiting perpendicular to the disk (Johnston et al., 1995; Ibata et al., 1997; Helmi & White, 2001; Newberg et al., 2002; Majewski et al., 2003). Since this discovery, other streams around the MW have been discovered such as the Helmi stream (Helmi et al., 1999; Chiba & Beers, 2000),

Palomar 5 (Odenkirchen et al., 2001), Cetus (Newberg et al., 2009), GD-1 (Grillmair & Dionatos, 2006), and many more. Because stellar streams are a consequence of the hierarchical formation of galaxies, many studies discovered them around other MW-mass galaxies as well (for example Peng et al., 2002; Zucker et al., 2004). With the advancement of the new Southern Stellar Stream Spectroscopic Survey (S^5; Li et al., 2019), the discovery of such systems will continue to inform our understanding of stream formation and evolution (Li et al., 2022).

1.2. Satellite Galaxies

Not all of the lower mass galaxies in the vicinity of the MW merged away or were tidally disrupted into streams. Another strong indicator of the hierarchical galaxy formation paradigm are the numerous low-mass, gravitationally bound galaxies orbiting within the MW halo, which we call 'satellite galaxies'. Many of the brightest, massive satellite galaxies (stellar masses $M_{\rm star} \gtrsim 10^5\,{\rm M}_\odot$) were known early on, such as the LMC, SMC, Draco, or Fornax, but with the advancement of larger, more technologically sophisticated surveys like the Sloan Digital Sky Survey (SDSS; Blanton et al., 2017) and the Dark Energy Survey (DES Dark Energy Survey Collaboration et al., 2016), many more lower mass satellites were discovered.

The lowest mass galaxies we know to-date are called 'ultra-faint' galaxies, which have stellar masses down to $M_{\rm star} \sim 10^{2-4}\,{\rm M}_\odot$, and dark matter halo masses of $M_{\rm halo} \sim 10^{8-9}\,{\rm M}_\odot$ (Bullock & Boylan-Kolchin, 2017; Simon, 2019), which are likely near the limits of galaxy formation. Another recently discovered galaxy, Crater II, has been classified as a 'low surface brightness' galaxy, given that its stellar mass is $M_{\rm star} \sim 3.5 \times 10^5\,{\rm M}_\odot$ but its physical size is roughly $\sim 1\,$kpc (Torrealba et al., 2019; Ji et al., 2021). Because their stars are spread over such a wide area on the sky, these galaxies are exceptionally difficult to find, however, others exist, such as the recently discovered Antlia II, with a stellar mass of $M_{\rm star} \sim 1.6 \times 10^6\,{\rm M}_\odot$ and radius of $\sim 2.5\,$kpc, comparable to the size of the LMC on the sky (Torrealba et al., 2016; Ji et al., 2021).

Today, we have measured many properties, including distances, velocities, magnitudes, and masses of $50 - 60$ satellites of the MW, and ~ 30 satellites of Andromeda (M31) (McConnachie,

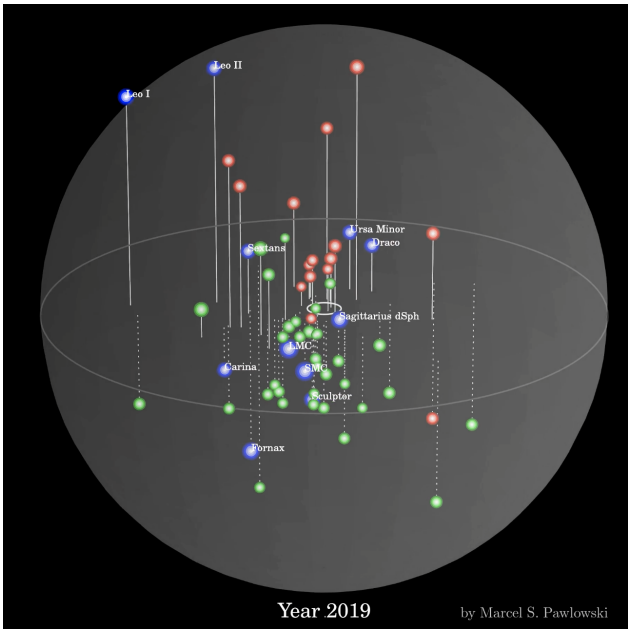

FIGURE 1.1. Distribution of satellite galaxies around the MW as of 2019. The disk of the MW is denoted by the small white ellipse in the center of the figure, while the colored points represent the satellite galaxies; blue points indicate the most massive 'classical' satellites, red points are satellites discovered through SDSS, and green points are the least massive satellites discovered from DES (Bullock & Boylan-Kolchin, 2017, Pawlowski).

2012; Fritz et al., 2018a; Putman et al., 2021). Figure1.1 shows the spatial distribution of the MW satellites as of 2019 (Bullock & Boylan-Kolchin, 2017).

1.2.1. Satellite Infall. For a satellite galaxy, one of the most significant events in its orbital history is when it first orbited within the dark matter halo of a more massive host galaxy. Satellites generally remain gravitationally bound to the host galaxy, although it is possible that they temporarily orbit outside of the halo and we refer to satellites in this temporary phase as 'splashback' satellites (Ludlow et al., 2009; Wetzel et al., 2014). The splashback radius, sometimes called the 'turn-around' radius, was a theoretical prediction from the 'spherical collapse model' (Fillmore & Goldreich, 1984; Bertschinger, 1985), and later seen in dark matter only simulations where the density of dark matter density steeply drops off, often near the first apocenter of particles/subhalos (Vogelsberger et al., 2011; Diemer & Kravtsov, 2014; Adhikari et al., 2014). As satellites fall

into the more massive host, they interact with the hot halo gas which can quench the satellite's star formation through the ram-pressure stripping of its gas (Gunn & Gott, 1972; van den Bosch et al., 2008; Fillingham et al., 2019; Rodriguez Wimberly et al., 2019; Samuel et al., 2022b). This process can be exacerbated through subsequent infall events into the massive host, or through the infall into another intermediate-mass host, before falling into the main host, in a process called 'pre-processing' (D'Onghia & Lake, 2008; Li & Helmi, 2008; Wetzel et al., 2015; Deason et al., 2015; Jahn et al., 2022; Samuel et al., 2022b). Studying the infall histories of satellite galaxies through the use of simulations to make sense of the MW's satellite population has been the focus of numerous works (for example, Slater & Bell, 2013; Wetzel et al., 2015; Li et al., 2020b; Bakels et al., 2021; D'Souza & Bell, 2021; Robles & Bullock, 2021; Ogiya et al., 2021; Rodriguez Wimberly et al., 2022).

1.2.2. Satellites of Satellites. The hierarchical build up of galaxies over time also suggests that lower mass galaxies could also host their own satellite populations. The LMC is the most massive satellite galaxy around the MW with a stellar mass of $M_{\rm star} \sim 2 \times 10^9\,{\rm M}_\odot$, or about $1/20$ the stellar mass of the MW (van der Marel et al., 2002; McConnachie, 2012). The LMC also has it's own satellite population including the SMC, $M_{\rm star} \sim 4-5\times10^8\,{\rm M}_\odot$, and numerous studies have focused on determining which of the MW's satellites are associated with the LMC through observational evidence and simulations (Deason et al., 2015; Jethwa et al., 2016; Kallivayalil et al., 2018; Jahn et al., 2019; Pardy et al., 2020; Patel et al., 2020). Similarly, the massive M33 satellite of M31, $M_{\rm star} \sim 3 \times 10^9\,{\rm M}_\odot$ (McConnachie, 2012), hosts its own satellites, but because of its distance to us it is much more difficult to find them (Martínez-Delgado et al., 2022). Because the LMC (and M33) hosts its own satellites, shortly after falling into the MW they should all have roughly similar orbits and qualitatively similar orbital energies and angular momenta (Sales et al., 2011; Deason et al., 2015; Sales et al., 2017; Pardy et al., 2020; Patel et al., 2020). There is also increasing evidence that the LMC is on its first infall into the MW and near its first pericentric passage today (Kallivayalil et al., 2006; Besla et al., 2007; Kallivayalil et al., 2013; Patel et al., 2020).

1.2.3. Orbital Dynamics. Characterizing the dynamics of satellite populations around MW-mass galaxies, and quantifying their orbit properties, gives us insight to their co-evolution with the host. As such, measuring their 6D phase space coordinates is crucial for determining their orbits in the past. We measure galaxy distances by observing stars on the tip of the red giant branch which have a unique, known luminosity that we calibrate to (Da Costa & Armandroff, 1990; Lee et al., 1993), or variable stars which have a relationship of luminosity and rate of variability (Leavitt, 1908). We measure the radial velocities of galaxies by measuring the redshift of their spectral lines. The other component of the velocity is the proper motion, which is the apparent motion of a galaxy across the sky. However, proper motions are difficult to measure precisely because they require at least two measurements, and more to reduce the uncertainties. Furthermore, the farther a galaxy is from us, the longer we need to wait between observations to measure an appreciable difference in the position of the galaxy.

Numerous studies, and recent Hubble Space Telescope (HST) treasury programs (GO-14734, PI Kallivayalil; GO-15902, PI Weisz), have enabled us to obtain full kinematic measurements, including proper motions, for all satellites in the LG (for example McConnachie, 2012; Kallivayalil et al., 2013, 2018; Fritz et al., 2018a,b; Gaia Collaboration et al., 2018; Patel et al., 2020). Excitingly, many studies are now using a combination of HST and Gaia data in conjunction to decrease the uncertainty in satellite galaxy proper motions (del Pino et al., 2022; Bennet et al., 2022; Warfield et al., 2023), and similar studies will make use of upcoming James Webb Space Telescope (JWST) data (Weisz et al., 2023). The Satellites Around Galactic Analogs (SAGA) survey (Geha et al., 2017; Mao et al., 2021) is also studying the satellites around other MW-mass galaxies. Thus, understanding the orbit and infall properties of the satellites in the LG is important to give context to future results.

1.3. Modeling Satellite Galaxy Orbits

To model the orbits of satellite galaxies, of equal importance to their 6D phase space coordinates, is the mass distribution, or global gravitational potential of the MW. Some studies use analytic

descriptions of the various components of the MW, like the bulge, thin disk, thick disk, and dark matter halo, and fit observational data to these models to derive parameters that best describe the mass distribution of the MW (for example McMillan, 2011; Bovy & Rix, 2013; Bovy et al., 2016a; McMillan, 2017). Other studies used Bayesian modeling or cosmological simulations in conjunction with the kinematic properties of globular clusters or satellite galaxies to estimate the total mass of the MW or M31 (Eadie et al., 2017; Patel et al., 2018; Eadie & Jurić, 2019; Patel et al., 2020).

Although a number of works give us decent approximations of the mass distribution of the MW, we do not know exactly how it has changed over time. Thus, a number of studies keep the mass or potential of the MW fixed, while backward integrating the orbits of satellites to estimate their orbit histories (Patel et al., 2017; Fritz et al., 2018a; Fillingham et al., 2019; Pace et al., 2022). However, given hierarchical formation, we know that this model is incorrect. Using these static, fixed potentials, some authors factor in more complexity to the model by including the gravitational contribution of the LMC (Garavito-Camargo et al., 2019; Patel et al., 2020; Garavito-Camargo et al., 2021; Samuel et al., 2021; Vasiliev et al., 2021; Correa Magnus & Vasiliev, 2022; D'Souza & Bell, 2022; Pace et al., 2022; Vasiliev, 2023), or by including dynamical friction (for example Weinberg, 1986; Patel et al., 2018; Garavito-Camargo et al., 2019, 2021) which acts to slow satellites down and ultimately robs them of their orbital energy or angular momentum. Nevertheless, understanding the limitations of all of these methods remains challenging (Klypin et al., 1999). To derive the orbits of satellite galaxies, or any gravitationally bound system, many of these studies rely on recently developed numerical tools, such as galpy (Bovy, 2015), AGAMA (Vasiliev, 2019a), and Gala (Price-Whelan, 2017), or simulations of galaxies like the MW, with or without the inclusion of baryonic physics.

1.4. Cosmological Simulations

1.4.1. Dark Matter-only Simulations. In recent years, technological and algorithmic advancements have made it possible to simulate galaxies in a cosmological context to better study

how galaxies form, and the physical processes important in shaping galaxy-wide properties, such as star formation rate, total stellar mass, density profiles, and rotation curves. Early N-body simulations only included dark matter particles and studied the growth and dynamical evolution of galaxies (Press & Schechter, 1974; White, 1976), the clustering of galaxies while testing different cosmologies (Gott et al., 1979; Colberg et al., 2000; Evrard et al., 2002), and more recently, the dark matter distribution and evolution on Mpc to Gpc scales (Springel et al., 2005) or the halo mass function (Boylan-Kolchin et al., 2009). With each passing decade, the number of dark matter particles used to model populations of galaxy halos, or galaxy clusters, increased from $O(10^2)$ in the earliest simulations, to nearly $O(10^{10})$ today. Because of the ever increasing number of particles, the numerical methods used to calculate the gravitational interactions between particles has also improved from first calculating forces from all possible particle pairs, to using 'tree'-like methods which groups particles together and approximates their combined effect on a given particle, based on their density and distance (Barnes & Hut, 1986; Kravtsov et al., 1997; Springel, 2005).

1.4.2. Modeling Baryonic Processes. Although the dark matter content in the Universe is $\approx 5\times$ that of the normal baryonic matter (Planck Collaboration et al., 2018), modeling baryonic processes that occur throughout galaxy formation and evolution is key to forming realistic galaxies (for example, see Sales et al., 2022, and references therein). Simulations that model these baryonic processes treat the numerics in different ways. 'Eulerian'-based methods solve the fluid equations on either a fixed grid/mesh, or allow the mesh to adapt to the local density at each snapshot, a process called Adaptive Mesh Refinement (AMR; Kravtsov et al., 1997; Teyssier, 2002; Bryan et al., 2014). 'Lagrangian'-based methods, in contrast, treat the gas and stars as individual cells/particles, and compute the hydrodynamic forces between them. Many Lagrangian codes use a common technique called 'Smoothed Particle Hydrodynamics' (SPH Springel, 2005, 2010a), where the particles are 'smoothed' out based on the number of nearest neighbors to it.

While the methods above work well in certain scenarios, they each come with their own disadvantages. For instance, AMR may result in decreased accuracy at the the mesh boundaries,

and SPH may suppress instabilities between fluids and not capture shock-fronts in detail(Springel, 2010b; Hopkins, 2015). To improve accuracy in hydrodynamic calculations, newer methods, such as 'moving-mesh' and 'meshless finite-mass/volume' (MFM/MFV), were developed to work around these limitations. Moving-mesh methods, such as AREPO, set up unstructured grids, called Voronoi tessellations, based on the distribution of mass at each timestep, and calculate the fluid equations while additionally accounting for the faces of the mesh to move (Springel, 2010b). This allows the resolution to be set by the density, similar to SPH, and better resolves fluid instabilities, similar to AMR. The MFM/MFV methods, on the other hand, are Lagrangian and use particles to simulate the fluid, rather than a mesh. As the fluid evolves, the resolution of the fluid adapts, while conserving mass, momentum, and energy to high precision (Hopkins, 2015). However, the MFM/MFV methods differ from traditional SPH based on how they estimate the equations of motions of particles, how they numerically calculate gradients, and the fact that the particles have a finite volume, and are not true point-particles (Hopkins, 2015).

Nevertheless, in order to properly simulate the formation of galaxies, a plethora of physical processes must be modeled, such as gas heating and cooling, star formation, stellar feedback processes such as stellar winds, radiation pressure, and supernovae (core collapse and Type Ia), black hole feedback from Active Galactic Nuclei, interactions of charged particles with magnetic fields, and how to properly model the effects and formation of dust. Every simulation group models each of these processes in their own distinct way, for better or worse, and still manage to produce realistic galaxies and galaxy clusters. As such, there are no 'best' set of models as of yet; see Vogelsberger et al. (2020) for a review of numerical techniques and baryonic processes.

1.4.3. Zooming in on a MW-mass galaxy. One area of focus in simulation-based studies is to zoom-in on a single galaxy within a cosmological context, and study or re-simulate it at higher resolution (Navarro & White, 1994; Hahn & Abel, 2011; Oñorbe et al., 2014). State-of-the-art simulation suites, such as the NIHAO (Buck et al., 2019, 2020), EAGLE (Schaye et al., 2015), APOSTLE (Sawala et al., 2016), or Auriga (Grand et al., 2017) simulations, among others simulate

galaxies similar to the MW, along with their satellite populations, with dark matter and gas particle masses between $m_{\rm DM} \sim 10^{4-7}\,{\rm M}_\odot$ and $m_{\rm gas} \sim 10^{4-6}\,{\rm M}_\odot$, respectively. Of equal importance to the mass resolution is the spatial resolution, which is often taken to be the minimum distance at which particles can interact with one another. This is sometimes called the force softening length and the simulations above resolve interactions between $\sim 100 - 1000$ parsec.

1.4.4. Feedback in Realistic Environments. In my work, I use the 14 MW-mass simulations from the Feedback In Realistic Environments (FIRE) collaboration. These simulations model the formation and evolution of a MW-mass galaxy, or MW/M31-like pairs of galaxies, across cosmic time in a cosmological context, with dark matter and gas particle resolutions of $m_{\rm DM} \approx 10^4$ and $m_{\rm gas} \approx 10^3$, respectively, dark matter spatial resolutions of $\sim 40\,{\rm pc}$, and gas spatial resolution which is adaptive down to $\sim 1\,{\rm pc}$ (Hopkins et al., 2014; Garrison-Kimmel et al., 2014; Hopkins, 2015; Wetzel et al., 2016; Hopkins et al., 2018; Garrison-Kimmel et al., 2019a). The host galaxies in this suite resemble MW-mass analogs (Garrison-Kimmel et al., 2018; El-Badry et al., 2018a; Hopkins et al., 2018), with disk morphologies, metallicity gradients, and dynamics (Ma et al., 2017; Sanderson et al., 2020; Bellardini et al., 2021, 2022; McCluskey et al., 2023), and stellar halos (Bonaca et al., 2017; Sanderson et al., 2018b) similar to the MW and other similar mass galaxies. Each MW-mass host possesses its own satellite population with realistic stellar masses and kinematics (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a; Santistevan et al., 2023), radial and 3D distributions (Samuel et al., 2020, 2021), and star formation histories (Garrison-Kimmel et al., 2019b).

1.5. Dissertation Outline

This dissertation consists of four studies on the evolution of low-mass galaxies across cosmic time. In Chapter 2, I will quantify and describe the low-mass galaxies that merged together over time to build a galaxy like the MW, and I will provide estimates of when MW-mass galaxies typically form based on different definitions of galaxy formation. Chapter 3 highlights the origins of a unique

subset of metal-poor stars on prograde orbits seen in the MW's disk, and simulations, today, which come from a gas-rich, LMC/SMC-mass merger roughly 7-12.5 Gyr ago. For the low-mass galaxies that survive to present-day, I will discuss when these galaxies first became satellites, and discuss trends in their instantaneous orbital dynamics, and orbit histories, with their infall time, present-day distance from the MW-mass host, and satellite stellar mass in Chapter 4. Finally, in Chapter 5 I will compare the orbit histories of the satellites in cosmological simulations against orbits derived in a fixed, axisymmetric MW-mass host potential model to highlight the strengths and limitations of this commonly applied technique.

CHAPTER 2

The Formation Times and Building Blocks of Milky Way-mass Galaxies in the FIRE Simulations

Published as Isaiah B. Santistevan, Andrew Wetzel, Kareem El-Badry, Joss Bland-Hawthorn, Michael Boylan-Kolchin, Jeremy Bailin, Claude-André Faucher-Giguère, Samantha Benincasa in Monthly Notices of the Royal Astronomical Society, Volume 497, Issue 1, September 2020, Pages 747-764, https://doi.org/10.1093/mnras/staa1923

2.1. Abstract

Surveys of the Milky Way (MW) and M31 enable detailed studies of stellar populations across ages and metallicities, with the goal of reconstructing formation histories across cosmic time. These surveys motivate key questions for galactic archaeology in a cosmological context: when did the main progenitor of a MW/M31-mass galaxy form, and what were the galactic building blocks that formed it? We investigate the formation times and progenitor galaxies of MW/M31-mass galaxies using the FIRE-2 cosmological simulations, including 6 isolated MW/M31-mass galaxies and 6 galaxies in Local Group (LG)-like pairs at $z = 0$. We examine main progenitor "formation" based on two metrics: (1) transition from primarily *ex-situ* to *in-situ* stellar mass growth and (2) mass dominance compared to other progenitors. We find that the main progenitor of a MW/M31-mass galaxy emerged typically at $z \sim 3-4$ ($11.6-12.2\,$Gyr ago), while stars in the bulge region (inner 2 kpc) at $z = 0$ formed primarily in a single main progenitor at $z \lesssim 5$ ($\lesssim 12.6\,$Gyr ago). Compared with isolated hosts, the main progenitors of LG-like paired hosts emerged significantly earlier ($\Delta z \sim 2$, $\Delta t \sim 1.6\,$Gyr), with $\sim 4\times$ higher stellar mass at all $z \gtrsim 4$ ($\gtrsim 12.2\,$Gyr ago). This highlights the importance of environment in MW/M31-mass galaxy formation, especially at

early times. On average, about 100 galaxies with $M_{\rm star} \gtrsim 10^5\,{\rm M}_\odot$ went into building a typical MW/M31-mass system. Thus, surviving satellites represent a highly incomplete census (by $\sim 5\times$) of the progenitor population.

2.2. Introduction

The properties of stellar populations within a galaxy, including their age, elemental abundances, and kinematics, all provide rich insight into the galaxy's formation history. For example, early studies of the kinematics of stars with highly radial orbits in the Milky Way (MW) suggested that these stars must have formed differently from those on more disk-like orbits, implying that the MW formed via a gravitational collapse (Eggen et al., 1962). Other early studies proposed that stars and clusters in the outer halo formed from material in proto-galaxies that continued to fall into the galaxy after the central regions already collapsed (e.g. Ostriker & Tremaine, 1975; Searle & Zinn, 1978). We know today that the processes involved in galaxy formation are more elaborate (Freeman & Bland-Hawthorn, 2002).

Initial theories for galaxy formation invoked the dissipational collapse of gas in dark-matter (DM) halos (e.g. Rees & Ostriker, 1977; White & Rees, 1978; Fall & Efstathiou, 1980; Mo et al., 1998). More recent works have examined the effects of galaxy mergers as well (e.g. Springel & Hernquist, 2005; Robertson et al., 2006; Stewart et al., 2008; Garrison-Kimmel et al., 2018). In gas-rich mergers, stellar feedback plays an important role in retaining gas content prior to the merger as it can heat the interstellar medium (ISM) and redistribute gas throughout the galaxy, even to larger radii where the effects of gravitational torquing are not as strong (Hopkins, 2009). Without feedback, the gas easily is torqued and falls to the center of the gravitational well, where it gets consumed in a starburst. This implies that some part of galactic disks must survive a merger process, and the thick disk of the MW likely survived a significant merging event, which could have deposited fresh gas into the MW and dynamically heated stars from the thick disk into the stellar halo (Gallart et al., 2019).

The stellar halo of the MW is perhaps the best place to probe the remnants of its early formation process. Various works have studied hierarchical formation of the stellar halo (e.g. Bullock et al., 2001; Bullock & Johnston, 2005; Helmi, 2008; Johnston et al., 2008; Cooper et al., 2010; Deason et al., 2016), showing that it occurs via the tidal disruption and accretion of many satellite dwarf galaxies. For instance, using cosmological zoom-in simulations, Deason et al. (2016) find that typically $1-2$ satellite galaxies contribute most of the accreted stellar material to a stellar halo. More generally, they find that the majority of accreted metal-poor stars come primarily from 'classical' dwarf galaxies ($\sim 40-80$ per cent) as opposed to 'ultra-faint' dwarf galaxies (only $\sim 2-5$ per cent). They also find a relation between the galaxy's progenitor mass and its satellite population at $z = 0$: galaxies with less massive progenitors tend to have more quiescent histories, as well as a less massive surviving satellite population, when compared to the more massive galaxies. Similarly, examining the AURIGA simulations, Monachesi et al. (2019) find a correlation with the number of 'significant progenitors' (number of progenitors that contribute 90 per cent of the stellar halo mass) and the accreted mass in the halo, with more massive halos accreting smaller numbers of significant progenitors. Studies of kinematically coherent structures in the MW's halo, like the Sagittarius stream (Newberg et al., 2002; Majewski et al., 2003) and Gaia-Enceladus (Helmi et al., 2018; Belokurov et al., 2018) clearly confirm this hierarchical formation scenario. In particular, Gaia-Enceladus is thought to be comparable in stellar mass to the SMC ($\sim 6\times10^8$ $M_\odot$), contributing most of the stars in the (inner) stellar halo (Helmi et al., 2018).

Studies of old and/or metal-poor stars provide the best window into the early formation of the MW (e.g. Brook et al., 2007; Scannapieco et al., 2006; Deason et al., 2016; Griffen et al., 2018; Sestito et al., 2019; Chiaki & Wise, 2019). Current spectroscopic surveys (e.g. RAVE, GALAH, APOGEE, LAMOST) now provide elemental abundances and ages for stars across the MW (e.g. Steinmetz et al., 2006; De Silva et al., 2015; Majewski et al., 2017; Li et al., 2015). These surveys achieve high spectral resolution (up to $R \sim 30,000$) and signal-to-noise (up to $S/N > 100$), and are capable of observing stars with $[\mathrm{Fe/H}] < -2$. Recently, the *Pristine* survey has observed significant

populations of stars at [Fe/H] < -3, with promise of reaching to [Fe/H] < -4 (Starkenburg et al., 2017b). Interestingly, using observations of metal-poor stars, Sestito et al. (2019) found that a significant fraction of these stars with [Fe/H] < -4 are on disk-like orbits in the MW. Similar work using metal-poor stars from LAMOST also has proven useful in finding halo structures (e.g. Yuan et al., 2019). A key question is where these metal-poor stars come from. Did they form within the MW, or did they form in other dwarf galaxies that subsequently merged in? If the latter, it would not make sense to say that they formed in the MW, or at least, its main progenitor.

Recently, using the FIRE-2 cosmological zoom-in baryonic simulations, El-Badry et al. (2018b) predicted that the oldest ($z_{\rm form} > 5$), metal-poor stars ([Fe/H] $\lesssim -2$) in the MW should be less centrally concentrated than stars that formed later, because (1) early merger events deposited stars that formed in dwarf galaxies on dispersion-supported orbits, and (2) stars that formed within the primary galaxy were heated to larger orbits via feedback-driven time-varying galactic potential. A similar study by Starkenburg et al. (2017a) found comparable results in the APOSTLE simulations of MW/M31-like pairs for stars with $z_{\rm form} > 6.9$ and [Fe/H] $\lesssim -2.5$.

While current MW surveys give us detailed information about a star's position, kinematics, and elemental abundances, obtaining precise ages for stars remains challenging. Current methods include fitting isochrones to stellar populations in colour-magnitude diagrams, studying oscillation modes of individual stars (astroseismology), using the rotation-age relation to infer ages (gyrochronology), and using detailed elemental abundances (e.g. Chaplin et al., 2014; Martig et al., 2016; Creevey et al., 2017; Silva Aguirre et al., 2018). Uncertainties in the ages of stars using the latter method can be as large as 40 per cent, but this improves if one can use a combination of methods (e.g. Miglio et al., 2017). Gaia's second data release (DR2; Gaia Collaboration et al., 2018) now provides distance measurements for over 1.3 billion stars in the MW, allowing astronomers to measure isochrone ages. For example, Gallart et al. (2019) suggest that the MW halo formed 50 per cent of its stars by $z \sim 4.2$ (12.3 Gyr ago) and the thick disk formed 50 per cent by $z \sim 2$ (10.5 Gyr ago). Other analyses reported formation lookback times of the halo of $8 - 13$ Gyr (e.g.

Schuster et al., 2012; Hawkins et al., 2014), including different formation times for different halo populations (e.g. Ge et al., 2016). Measurements of stellar ages in the MW bulge suggest that the stellar population is predominantly older than $\sim$ 10 Gyr (Barbuy et al., 2018, and references therein).

Cosmological galaxy simulations provide the best theoretical laboratories for understanding the full evolutionary histories of galaxies across cosmic time. It remains unclear if the formation histories of MW/M31-mass galaxies depend on whether they are in isolated environments (with no other nearby companions of similar mass) or in LG-like environments (with a pair of massive galaxies at $\lesssim$ 1 Mpc separation). Most simulations of MW/M31-mass galaxies, either idealized or cosmological focus on a *single isolated* galaxy/halo (e.g. AURIGA, NIHAO, ERIS, Caterpillar), not in a Local Group (LG)-like MW+M31 pair (Grand et al., 2017; Buck et al., 2019; Guedes et al., 2011; Griffen et al., 2016). Exceptions include the ELVIS DM-only (DMO) simulation suite, which include 24 MW/M31-mass halos in LG-like pairs and a mass-matched sample of 24 isolated halos (Garrison-Kimmel et al., 2014), the cosmological baryonic simulations of LG-like pairs from the APOSTLE suite (Sawala et al., 2016), and the simulated LG analogues from the CLUES project (Carlesi et al., 2020). Furthermore, semi-analytic models applied in a cosmological settings provide complementary tools to undersand MW-mass formation histories and their environmental dependence. For instance, the semi-analytical code GAlaxy MErger Tree and Evolution (GAMETE, and the similar code GAMESH which accounts for radiative transfer) allows for a detailed modeling for things such as the gas evolution within a MW-mass galaxy, the evolution of the galactic SFR, and the formation/evolution of early Population III (Pop III) stars (Salvadori et al., 2007; de Bennassuti et al., 2014; Graziani et al., 2015).

In addition to potential differences in host galaxy properties in LG-like versus isolated environments, it is also imperative to understand potential differences in their satellite populations. While observations of the LG have driven most of our knowledge of dwarf galaxy populations,

recent observational campaigns aim to measure satellite populations around (primarily more isolated) MW-mass galaxies such as M94 (Smercina et al., 2018), M101 (Danieli et al., 2017), M81 (Karachentsev & Kudrya, 2014), and Centaurus A (Müller et al., 2019). The Satellites Around Galactic Analogs (SAGA) survey[1] also is observing satellite populations around (mostly) isolated MW-mass galaxies down to the luminosity of Leo I ($M_r < -12.3$; $M_{star} \approx 5 \times 10^6\,M_\odot$), with a predicted sample size of 100 galaxies (Geha et al., 2017). Other campaigns instead focus on groups of galaxies out to $\sim$ 40 Mpc (e.g. Kourkchi & Tully, 2017). In connecting these observations with those of the LG, we must understand: *does environment play a role in the satellite galaxy populations, and thus building blocks, of MW/M31-mass galaxies across cosmic time?*

In this work, we use FIRE-2 simulations of 12 MW/M31-mass galaxies to investigate their cosmological hierarchical formation histories. We quantify these galaxies' building blocks across cosmic time and determine when their main progenitors formed/emerged. We investigate the times at which (1) the *stellar* mass growth of the most massive progenitor transitions from being dominated by ex-situ stars (via mergers) to in-situ star formation, and (2) when the most massive galaxy starts to dominate in stellar mass compared to other progenitor galaxies. When a progenitor galaxy satisfies either of these two criteria, we define it to be the 'main' progenitor, which is distinct from the other progenitor galaxies that continue to merge in. There are many definitions of host galaxy 'formation', such as when the main galaxy reaches a fraction of its mass at $z = 0$; here we refer to formation specifically as when a single main progenitor galaxy emerges in its environment, based both on its in-situ star formation (Section 2.4.3) and how its stellar mass compares to its neighbors (Section 2.4.4). We also emphasize that in our analysis, we focus on the formation of the host galaxy generally, and not specifically on the formation of a given component, such as the thin or thick disk.

Our work, however, is not the first to investigate general properties of the progenitors of MW-mass galaxies and their satellites. Many studies, including those from Dixon et al. (2018); Safarzadeh et al. (2018); Graziani et al. (2017); Magg et al. (2018); de Bennassuti et al. (2017),

[1]See the SAGA survey web site: http://sagasurvey.org

have focused on the high-z Universe to study star formation near the epoch of reionization, the metallicity distribution function, identifying ultra-faint dwarf galaxy progenitors, and Pop III stars to understand the early MW environment.

The main questions that we address are:

a) What were the building blocks (progenitor galaxies) of MW/M31-mass galaxies, and how many were there across cosmic time?

b) When did the main progenitor of a MW/M31-mass galaxy form/emerge?

c) Does the formation of MW/M31-mass galaxies depend on their environment, specifically, comparing isolated hosts to those in LG-like pairs?

2.3. Methods

TABLE 2.1. Properties of the 12 host galaxies in the FIRE-2 simulation suite that we analyze. Column list: name; stellar mass ($M_{star,90}$) within $R_{star,90}$; disk radius enclosing 90 per cent of the stellar mass within 20 kpc ($R_{star,90}$); halo virial mass (M_{200m}); halo virial radius (R_{200m}); redshift when the galaxy reached 50 per cent ($z_{0.5}$) or 10 per cent ($z_{0.1}$) of its stellar mass at $z = 0$; redshift when the cumulative fraction of stars that formed in-situ exceeded 0.5 when selecting stars at $z = 0$ within host-centric distances of 15 kpc ($z_{in-situ,15}$) and 2 kpc ($z_{in-situ,2}$; redshift when which the most massive progenitor exceeded a 3:1 stellar mass ratio with respect to the second most massive progenitor ($z_{MR,15}$); and the paper that introduced each simulation. Hosts with names starting with 'm12' are isolated hosts from the Latte suite, while the rest are in Local Group (LG)-like pairs from the ELVIS on FIRE suite.

Name	$M_{star,90}$ [10^{10} $M_\odot$]	$R_{star,90}$ [kpc]	M_{200m} [10^{12} $M_\odot$]	R_{200m} [kpc]	$z_{0.5}$	$z_{0.1}$	$z_{in-situ,15}$	$z_{in-situ,2}$	$z_{MR,15}$	Ref
m12m	10.0	11.6	1.6	371	0.6	1.5	1.9	2.1	1.5	A
Romulus	8.0	12.9	2.1	406	0.7	1.4	5.3	> 6.0	1.7	B
m12b	7.3	9.0	1.4	358	0.6	1.7	2.4	4.2	3.7	C
m12f	6.9	11.8	1.7	380	0.5	2.0	3.7	> 6.0	2.9	D
Thelma	6.3	11.2	1.4	358	0.4	1.1	1.7	1.7	4.4	C
Romeo	5.9	12.4	1.3	341	1.0	2.6	> 6.0	> 6.0	> 6.0	C
m12i	5.5	8.5	1.2	336	0.6	1.7	3.1	3.5	1.7	E
m12c	5.1	9.1	1.4	351	0.5	1.3	1.9	5.3	3.1	C
m12w	4.8	7.3	1.1	319	0.4	1.4	2.7	3.4	1.2	F
Remus	4.0	11.0	1.2	339	0.9	2.7	3.7	> 6.0	1.2	B
Juliet	3.3	8.1	1.1	321	0.8	2.4	5.3	5.5	4.7	C
Louise	2.3	11.2	1.2	333	0.9	2.5	4.9	5.0	3.5	C

Note: The references for each host are: A: Hopkins et al. (2018), B: Garrison-Kimmel et al. (2019a), C: Garrison-Kimmel et al. (2019b), D: Garrison-Kimmel et al. (2017b), E: Wetzel et al. (2016), and F: Samuel et al. (2020).

2.3.1. FIRE-2 Simulations of Milky Way- and M31-mass Galaxies. We use cosmological zoom-in baryonic simulations of MW/M31-mass galaxies from the Feedback In Realistic Environments (FIRE) project[2] (Hopkins et al., 2018). We ran these simulations using the `Gizmo` N-body gravitational plus hydrodynamics code (Hopkins, 2015), with the mesh-free finite-mass (MFM) hydrodynamics method and the FIRE-2 physics model (Hopkins et al., 2018). FIRE-2 includes several radiative cooling and heating processes for gas such as free-free emission, photoionization/recombination, Compton scattering, photoelectric, metal-line, molecular, fine-structure, dust-collisional, and cosmic-ray heating across a temperature range of $10 - 10^{10}$ K. This includes the spatially uniform, redshift-dependent cosmic UV background from Faucher-Giguère et al. (2009), for which HI reionization occurs at $z_{\rm reion} \sim 10$. The simulations self-consistently generate and track 11 elemental abundances (H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe), including sub-grid diffusion of these abundances in gas via turbulence (Hopkins, 2016; Su et al., 2017; Escala et al., 2018).

Stars form from gas that is self-gravitating, Jeans unstable, molecular (following Krumholz & Gnedin, 2011), and dense ($n_H > 1000$ cm^{-3}). Once a star particle forms, inheriting mass and elemental abundances from its progenitor gas element, it represents a single stellar population, assuming a Kroupa (2001) initial mass function, and it evolves along stellar population models from `STARBURST99 v7.0` (Leitherer et al., 1999). FIRE-2 simulations include several different feedback processes, including core-collapse and Ia supernovae, mass loss from stellar winds, and radiation, including radiation pressure, photoionization, and photo-electric heating.

We generated cosmological zoom-in initial conditions for each simulation at $z \approx 99$, embedded within periodic cosmological boxes of lengths $70.4 - 172$ Mpc using the code `MUSIC` (Hahn & Abel, 2011). We saved 600 snapshots down to $z = 0$, which are spaced every ≈ 25 Myr. All simulations assume flat ΛCDM cosmology with parameters consistent with Planck Collaboration et al. (2018). Specifically, the Latte suite (excluding m12w) used $\Omega_{\rm m} = 0.272$, $\Omega_{\rm b} = 0.0455$, $\sigma_8 = 0.807$, $n_{\rm s} = 0.961$, $h = 0.702$. Thelma & Louise and Romulus & Remus both used the same cosmology as in the original ELVIS DMO suite: $\Omega_{\rm m} = 0.266$, $\Omega_{\rm b} = 0.0449$, $\sigma_8 = 0.801$,

[2]See the FIRE project web site: http://fire.northwestern.edu

$n_s = 0.963$, $h = 0.71$. Finally, Romeo & Juliet and m12w both used $\Omega_m = 0.31$, $\Omega_b = 0.048$, $\sigma_8 = 0.82$, $n_s = 0.97$, $h = 0.68$.

In this work, we analyze 12 MW/M31-mass galaxies; Table 2.1 lists their properties. 6 galaxies are from the Latte suite of isolated MW/M31-mass galaxies, introduced in Wetzel et al. (2016). We selected these halos with the following two criteria: (1) $M_{200m} = 1 - 2 \times 10^{12}\,M_\odot$ (e.g. Bland-Hawthorn & Gerhard, 2016), and (2) no similar-mass halo within $5 \times R_{200m}$ (for computational efficiency). Here, M_{200m} refers to the total mass within a radius, R_{200m}, containing 200 times the mean matter density of the Universe. We chose galaxy m12w with one additional criterion: having an LMC-mass satellite at $z \sim 0$ in the pilot DM-only simulation (see Samuel et al., 2020). The Latte simulations have gas and *initial* star particle masses of $7100\,M_\odot$, although because of stellar mass loss, the typical star particle at $z = 0$ is $\sim 5000\,M_\odot$. DM particles have masses of $3.5 \times 10^4\,M_\odot$. Gas elements use fully adaptive force softening, equal to their hydrodynamic smoothing, that adapts down to 1 pc. The gravitational softening lengths for star and DM particle are fixed at 4 and 40 pc (Plummer equivalent), comoving at $z > 9$ and physical thereafter.

We also use 6 galaxies from the "ELVIS on FIRE" suite of 3 Local Group (LG)-like MW+M31 pairs (Garrison-Kimmel et al., 2019a,b), which were selected with the following criteria: (1) two neighboring halos each with mass $M_{200m} = 1-3\times10^{12}\,M_\odot$, (2) total LG mass of $2-5\times10^{12}\,M_\odot$, (3) center separation of $600-1,000$ kpc at $z = 0$, (4) radial velocities of $v_{rad} < 0\,km\,s^{-1}$ at $z = 0$, and (5) no other massive halos within 2.8 Mpc of either host center. The ELVIS on FIRE simulations have $\approx 2\times$ better mass resolution than Latte, with initial star/gas particle masses of $3,500 - 4,000\,M_\odot$. Hopkins et al. (2018) examined the effect of mass resolution on the formation histories of both m12i and m12m, finding differences $\lesssim 10 - 20$ per cent in their star formation histories (SFHs) comparing star particle masses of $7,000\,M_\odot$ to $56,000\,M_\odot$. We thus expect that the factor of $\approx 2\times$ resolution difference between the Latte and ELVIS suites should not significantly affect their SFHs.

We simulated these galaxies using the cosmological zoom-in technique (Oñorbe et al., 2014): see Wetzel et al. (2016), Hopkins et al. (2018), and Garrison-Kimmel et al. (2019a) for more

detail. We emphasize that we selected these halos solely using the parameters above, with no prior on their formation/merger histories or satellite populations (other than m12w). Furthermore, the numerical implementation that we used to generate these simulations, the `Gizmo` source code and FIRE-2 physics model, were the same across all simulations. Thus, these 12 hosts should reflect random/typical samplings of MW/M31-mass formation histories within their mass and environmental selection criteria. However, as we show below, LG-like versus isolated host selection does lead to systematically different formation histories.

Both the Latte and ELVIS on FIRE simulation suites form MW/M31-mass galaxies with realistic populations of satellite galaxies, in terms of their stellar masses and velocity dispersions (dynamical masses) (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a), radial distributions (Samuel et al., 2020), and star-formation histories (Garrison-Kimmel et al., 2019b). The MW/M31-mass host galaxies in the simulations also show a range of morphologies (Garrison-Kimmel et al., 2018; El-Badry et al., 2018a), with properties that broadly agree with the MW and M31, such as the stellar-to-halo mass relation (Hopkins et al., 2018), disk structure & gas mass (Sanderson et al., 2020), age & metallicity gradients (Ma et al., 2017), and stellar halos (Bonaca et al., 2017; Sanderson et al., 2018b). The papers above provide detailed properties for at least some of the hosts. For example, Ma et al. (2017) investigated age and metallicity gradients, as well as the disk structure, of m12i. Sanderson et al. (2018b) presented surface densities of both the disk and bulge, and halo stellar masses for *all* simulations used in this work. Sanderson et al. (2020) lists the gas masses, disk scale heights, and SFR (at $z = 0$) for m12f, m12i, and m12m. Several upcoming works will present gas and stellar metallicities (Bellardini et al., in prep), and the evolution of SFR across redshift (Yu et al., in prep) for all of these simulated hosts.

2.3.2. Identifying Halos and Galaxies. We use the `ROCKSTAR` 6-D halo finder (Behroozi et al., 2013a) to identify DM (sub)halos, using $M_{200\mathrm{m}}$ as our halo definition. We generate a halo catalog using only DM particles at each of the 600 snapshots and use `CONSISTENT-TREES` (Behroozi et al., 2013b) to construct merger trees. Given the conservatively large zoom-in volume that we generate

for each host, all of the halos that we examine have zero contamination by low-resolution DM particles.

We briefly summarize the method that we use to assign star particles to halos in post-processing; see Necib et al. (2018) and Samuel et al. (2020) for details. Given each (sub)halo's radius and maximum circular velocity, $R_{\rm halo}$ and $V_{\rm circ,max}$, as returned by `ROCKSTAR`, we identify star particles whose positions lie within 0.8 $R_{\rm halo}$ (out to a maximum distance of 30 kpc) and whose velocities are within 2 $V_{\rm circ,max}$ of the (sub)halo's center-of-mass velocity. After this, we keep star particles if they meet the following two criteria. First, their positions are within 1.5 $\rm R_{star,90}$ (the radius that encloses 90 per cent of the stellar mass) of both the center-of-mass position of current member star particles and the halo center. This criterion ensures that the galaxy's center of mass coincides with the halo's center of mass. Second, their velocities are within 2 $\sigma_{\rm vel,star}$, the velocity dispersion of current member star particles, of the center-of-mass velocity of member star particles. We iterate these two criteria until the stellar mass converges to within 1 per cent. We then keep halos with at least 6 star particles, stellar density $> 300\,\rm M_\odot\,kpc^{-3}$ (at $\rm R_{50}$, the radius which encloses 50 per cent of the stellar mass), and halo bound mass fraction > 40 per cent. Henceforth, when we refer to a galaxy's stellar mass, we mean the mass that we calculate from this process. We checked that none of our results change significantly if we use other ways of measuring stellar mass, such as the mass within $\rm R_{star,90}$.

Our software for reading and analyzing halo catalogs, including assigning star particles, is available via the `HaloAnalysis` package[3], and our software for reading and analyzing particles from Gizmo snapshots is available via the `GizmoAnalysis` package[4]; we first developed and used these packages in Wetzel et al. (2016).

2.3.3. Selecting Progenitor Galaxies. In our analysis, we impose two additional criteria to select galaxies of interest. First, we examine only galaxies with stellar mass $M_{\rm star} \geq 10^5\,\rm M_\odot$, which corresponds to $\geq 14 - 29$ star particles, depending on simulation resolution. Second, we

[3]https://bitbucket.org/awetzel/halo_analysis
[4]https://bitbucket.org/awetzel/gizmo_analysis

analyze only galaxies that are progenitors of each MW/M31-mass system at $z = 0$, selecting stars at various host-centric distances at $z = 0$ to probe the formation histories of different regions of the host galaxy/halo. Specifically, we select progenitor galaxies that contribute star particles to the following spherical distances, d, with respect to the center of each MW/M31-mass galaxy at $z = 0$:

a) $d(z = 0) < 300\,\mathrm{kpc}$ (hereon d_{300}), corresponding to the entire host halo system (virial region), including the entire host galaxy, stellar halo, and surviving satellite galaxies;

b) $d(z = 0) < 15\,\mathrm{kpc}$ (d_{15}), corresponding to the entire host galaxy (bulge and disk) plus inner stellar halo;

c) $d(z = 0) < 2\,\mathrm{kpc}$ (d_2), corresponding to an inner bulge region.

While these represent relatively simple spherical distance selections, we also investigated a 'disk' selection, by selecting stars at $z = 0$ within $R = 4 - 15\,\mathrm{kpc}$ and $|Z| < 2\,\mathrm{kpc}$ (in cylindrical coordinates), as well as requiring stars to be on co-rotating disk-like circular orbits via Toomre diagram selection. While this reduces the overall amount of accreted stars (as expected), given our method to select progenitor galaxies (see next paragraph), this selection generally did not lead to significant differences in our results compared to d_{15}.

Having defined the star particles in each region at $z = 0$, to select progenitor galaxies at a given redshift, we compute how much stellar mass a galaxy at a given redshift contributed to these host-centric distances at $z = 0$, relative to the galaxy's total stellar mass at that redshift. We define this as the 'contribution fraction'. To be a progenitor, we require that a galaxy has a contribution fraction greater than 1 per cent, that is, at least 1 per cent of its mass (at a given redshift) ends up within the host-centric region. We checked how our metrics for progenitor formation (described in Sections 2.4.3 and 2.4.4, and summarized in 2.4.5) changed as we varied this contribution fraction requirement. For d_{300}, our results are not sensitive to it, because all contribution fractions are near 100 per cent, indicating that galaxies that contribute to the host halo contribute essentially all of their stars. For d_2, and to a lesser extent d_{15}, the contribution fractions are more broadly spread throughout 0 - 100 per cent, because stars from infalling galaxies get deposited across a range of d

by $z = 0$. We thus use 1 per cent contribution fraction as a conservative minimum, though we note that increasing this minimum would decrease the number of progenitors to the inner galaxy.

2.4. Results

To provide visual context to our analysis, Fig. 2.1 shows synthetic Hubble Space Telescope u/g/r composite stellar images of two of our isolated hosts—m12f (top) and m12m (bottom)—using `STARBURST99` to determine the SED for each star particle (given its age and metallicity) and following the ray-tracing methods in Hopkins et al. (2005, 2018) for an observer 1 Mpc away. These images do not include dust attenuation. Each star particle is represented as a spherical cloud, with size representative of the local density of star particles. These images show stellar luminosity along a given line of sight using a logarithmic color scale, with visible stellar densities ranging from $10^{-9} - 3 \times 10^{-2}$ $M_\odot$ pc^{-3}.

Left to right, each panel shows different times throughout the formation of each MW/M31-mass galaxy, including all stars out to 300 kpc (physical) from the most massive progenitor. At $z = 0$ (right), the MW/M31-mass host galaxy is clearly dominant, including 16-27 resolved surviving satellite galaxies and an extended stellar halo with stellar streams from disrupted satellites. Some of the visible streams formed via outflows of gas (Yu et al., 2020), while others form via satellite disruption. However, at $z = 5$ (left), the system was composed of a collection of many similar-mass galaxies that eventually merge together. Because many galaxies had similar mass and none of them alone dominated the stellar mass growth, there was no meaningful single 'main' progenitor at this time. The middle panel shows each host when its main progenitor 'formed/emerged', which we define as when the cumulative stellar mass was dominated by in-situ formation in a single progenitor (see Section 2.4.3). We define in-situ formation as being star formation which occurs in the most massive progenitor (MMP); we discuss details of how we calculate the *in-situ fraction* in the beginning of Section 2.4.3. Similarly, ex-situ formation is defined as that which occurs in any other progenitor galaxy; we discuss details about how we determine the *ex-situ fraction* in

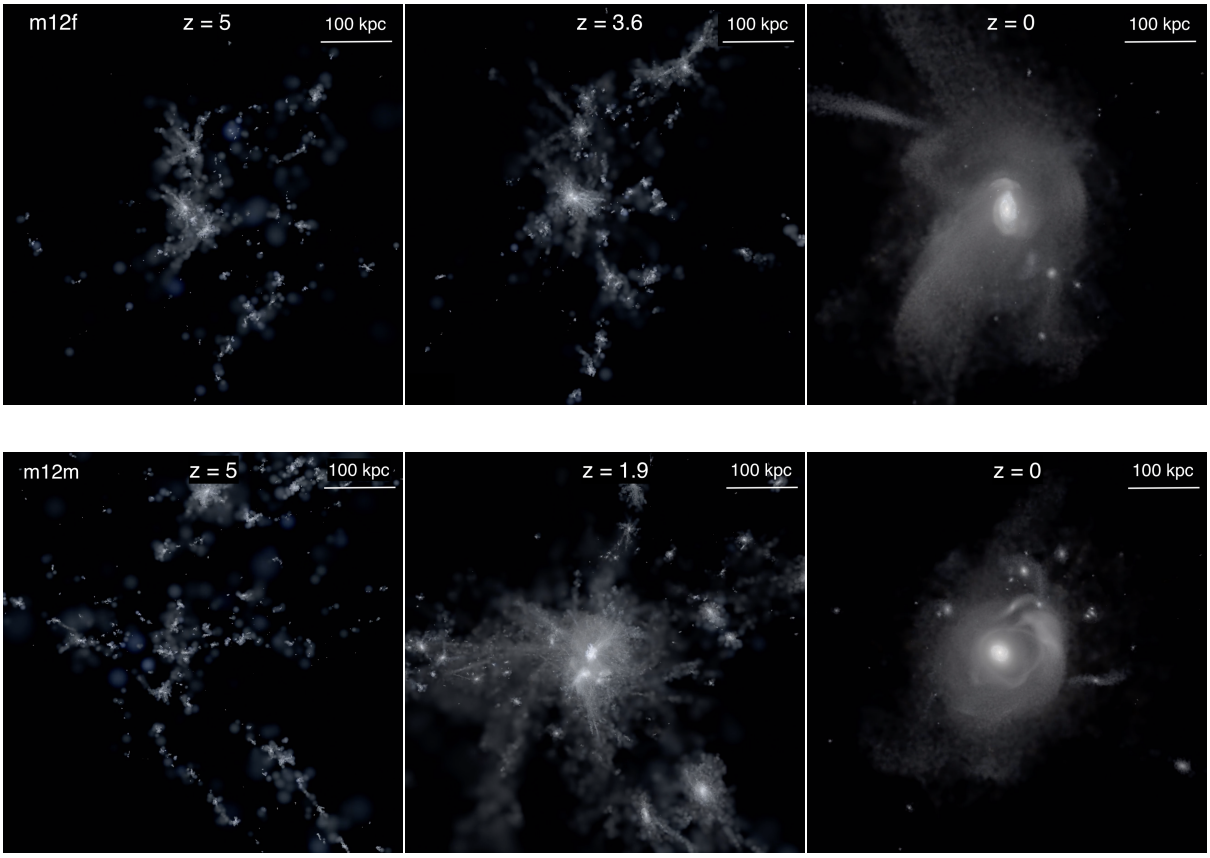

FIGURE 2.1. Synthetic Hubble Space Telescope u/g/r composite stellar images of all star particles (ignoring dust attenuation) out to a distance of 300 kpc physical (roughly $R_{200\mathrm{m}}$ at $z = 0$) around the most massive progenitor (MMP) at each redshift, for two simulations that bracket the main progenitor formation times across our isolated hosts. Each star particle is represented as a spherical cloud, with size representative of the local density of star particles. Intensity is logarithmic with stellar density, the visible regions ranging from $10^{-9} - 3 \times 10^{-2}$ $\mathrm{M}_\odot$ pc^{-3}. Left to right, the panels show images at $z = 5$, $z = z_{\mathrm{form}}$ (when the main progenitor 'formed' or 'emerged'), and at $z = 0$. We define main progenitor formation when the majority of stars for d_{15} have formed in-situ in a single progenitor (see Section 2.4.3); this is similar to when the stellar mass ratio of the most massive to second most massive progenitor exceeds 3:1 (see Section 2.4.4). The top row shows m12f, our earliest forming isolated host, and the bottom row shows m12m, our latest forming isolated host. *Prior to z_{form}, there was little meaningful sense of a single 'main' progenitor; instead, there was a collection of similar-mass progenitor galaxies.*

Section 2.4.2. Compared to the panels on the left, we see the emergence of a main progenitor galaxy that dominates its local environment in mass by more than 3:1 (see Section 2.4.4).

2.4.1. What was the mass growth of the most massive progenitor and does it depend on environment? First, we examine the stellar mass growth history of the MMP of each MW/M31-mass galaxy from $z = 0$ back to $z = 7$. Following the criteria above, we identify the progenitor galaxy with the highest stellar mass at each snapshot (according to our halo catalog) and label it as the MMP. This is different from using the halo merger tree to track the galaxy back in time, because DM halo and stellar masses grow at different rates, so the most massive halo is not always the most massive galaxy, especially at early times before a main progenitor has emerged. Furthermore, as we argue in Section 2.4.3, it does not make sense to think about the MMP as the single 'main' progenitor before the 'formation' redshift of each host. Our goal here is to set the stage by providing the relevant mass scale of the MMP across cosmic time while highlighting environmental differences. Also, the results in this subsection do not depend on the details of host-centric distance selection at $z = 0$, because the MMP contributes significant mass inside all of our distance cuts.

Figure 2.2 caption: **Top Left:** The stellar mass of the most massive progenitor (MMP) of each MW/M31-mass galaxy as a function of redshift (bottom axis) or lookback time (top axis). Thin colored lines show each simulation, including the 6 isolated hosts (solid) and 6 LG-like hosts (dotted). We also show the median $M_{\rm star}$ across all 12 hosts (thick solid black), and the median for the isolated (thick solid grey), and the LG-like (thick dashed grey) hosts. At $z = 0$ the host galaxies span a relatively narrow range of $M_{\rm star}(z = 0) = 2.3 - 10 \times 10^{10}\,{\rm M}_\odot$ (by selection), but because of scatter in formation history, their MMPs spanned about 1-1.5 orders of magnitude at all $z \gtrsim 4$ (> 12.2 Gyr ago), with a range of $6 \times 10^6 - 7 \times 10^7\,{\rm M}_\odot$ at $z = 7$, near the end of cosmic reionization. **Top Right:** Same as left but normalized to the median $M_{\rm star}(z)$ across all simulations. Romeo reached the highest value of ~ 6.5 at $z = 4 - 5$ ($12.2 - 12.6$ Gyr ago), while m12i was the lowest at ~ 0.15. *The MMPs of LG-like paired hosts had significantly higher mass on average, by up to a factor of 6, at all $z \gtrsim 2$.* **Bottom Left:** Same as top left but showing *fractional* mass growth, with each MMP normalized by its $M_{\rm star}(z = 0)$. The dotted horizontal lines indicate 10 per cent and 50 per cent of final mass. Considering the overall median, the MMP reached 10 per cent of

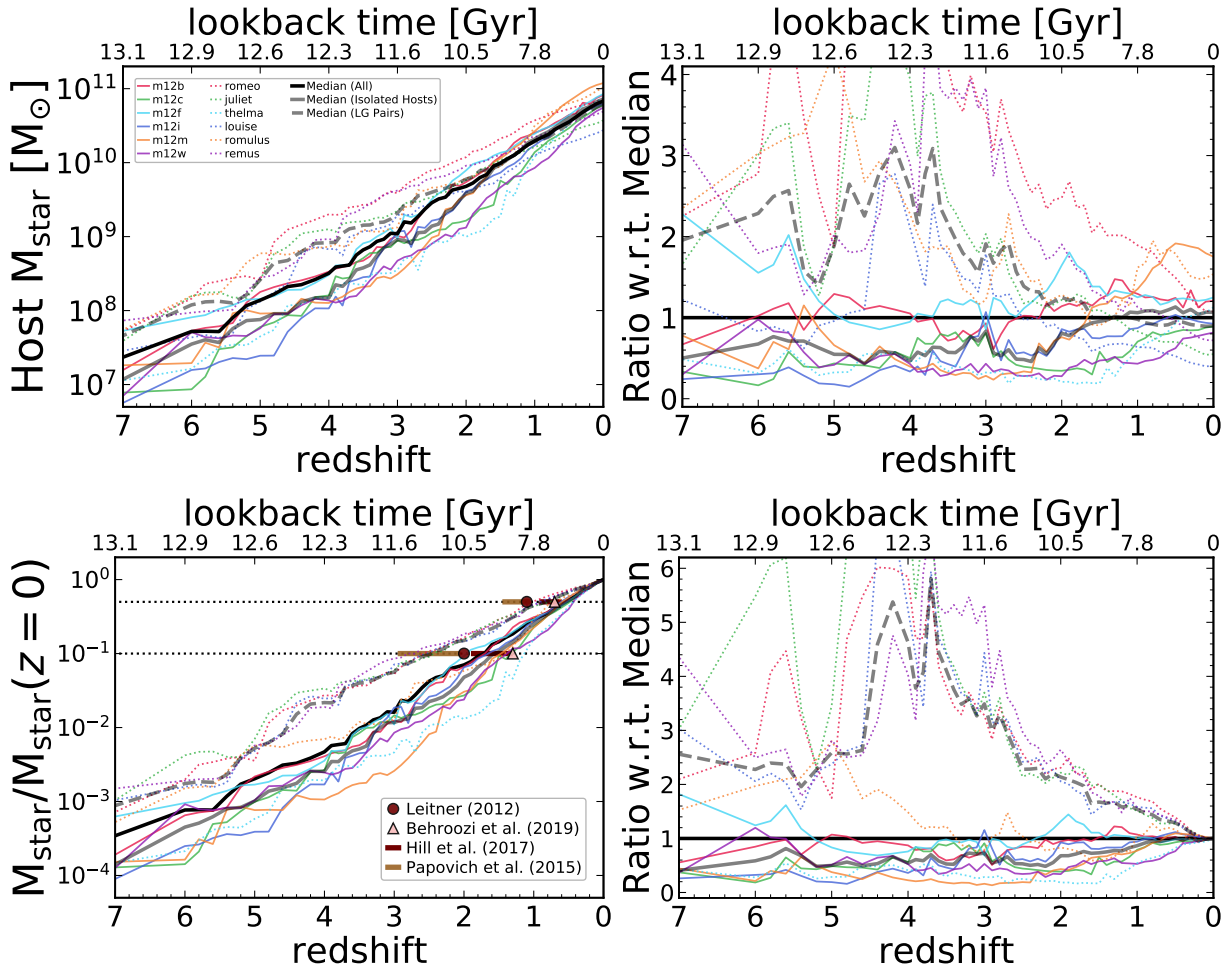

FIGURE 2.2. See the figure caption in the text.

its final mass by $z \sim 1.7$ (9.9 Gyr ago), while for isolated hosts it was $z \sim 1.5$ (9.4 Gyr ago) and for LG-like pairs it was $z \sim 2.4$ (11.0 Gyr ago). The MMP reached 50 per cent of its final mass by $z \sim 0.5$ (5.1 Gyr ago) across all hosts (and isolated host), and while this occurred at $z \sim 0.8$ (6.9 Gyr ago) for LG-like hosts. We also plot observational inferences for galaxies that span our mass range at $z = 0$ (see text). **Bottom Right:** Same as bottom left, but normalized to the total median: $[M_{\rm star}(z)/M_{\rm star}(z=0)]\,/\,[M_{\rm star}(z)/M_{\rm star}(z=0)]_{\rm med}$. Here, the enhancement for LG-like hosts is even most dramatic, *persisting at all redshifts and being almost* $10\times$ *that of isolated hosts at* $z \sim 4.2$ (12.3 Gyr ago). This highlights the importance of environment in MW/M31-mass galaxy formation, especially at early times.

Fig. 2.2 (top left) shows the stellar mass of the MMP versus redshift (bottom axis) and lookback time (top axis). We show each simulation individually (colored lines), including 6 isolated hosts (solid) and 6 LG-like hosts (dotted). We also show the median across all 12 hosts (thick solid black), and the median for the isolated (thick solid grey) and the LG-like (thick dashed grey) hosts. For clarity, we define $M_{\rm star}(z)$ as the stellar mass inside of the MMP galaxy that has formed up to redshift z. At $z = 0$ the host galaxies span a relatively narrow range of $M_{\rm star}(z = 0) = 2.3 - 10 \times 10^{10}$ M$_\odot$ (by selection). The shapes of the MMP mass growth histories show broad similarities, with $\log M_{\rm star}$ growing almost linearly with redshift. However, because of scatter in formation history, the $M_{\rm star}$ of MMPs spanned 1-1.5 orders of magnitude at $z \gtrsim 2$ (> 10.5 Gyr ago). For example, the MMPs spanned a range of $5.6 \times 10^6 - 7.3 \times 10^7$ M$_\odot$ at $z = 7$, near the end of cosmic reionization.

Fig. 2.2 (top right) shows the same, except we normalize each MMP to the overall median $M_{\rm star}(z)$, which more clearly highlights both scatter in mass growth and systematic differences between LG-like and isolated hosts. In particular, the typical $M_{\rm star}$ of the MMP of an isolated host is similar (slightly higher) than that of a LG-like host at all $z < 1.5$, but prior to this, the MMP of an LG-like host was significantly more massive than that of an isolated host. The biggest difference occurred at $z \sim 4$ (~ 12.2 Gyr ago), when LG-like hosts were $\sim 6\times$ more massive than isolated hosts.

Fig. 2.2 (bottom left) shows the *fractional* $M_{\rm star}$ growth of each MMP, that is, $M_{\rm star}(z)/M_{\rm star}(z = 0)$. This metric is more fair to compare the formation histories, because it normalizes out the scatter in $M_{\rm star}(z = 0)$. Across all simulations, the MMP reached 10 per cent and 50 per cent of its final mass at typically $z = 1.7$ (9.9 Gyr ago) and $z = 0.5$ (5.1 Gyr ago), respectively. Again, the MMP of an isolated host typically formed later, reaching 10 per cent and 50 per cent of its final mass at $z = 1.5$ (9.4 Gyr ago) and $z = 0.5$ (5.1 Gyr ago), while the MMP of a LG-like host typically formed earlier, reaching 10 per cent and 50 per cent of its final mass at $z = 2.4$ (11.0 Gyr ago) and $z = 0.8$ (6.9 Gyr ago).

Finally, Fig. 2.2 (bottom right) shows the fractional $M_{\rm star}$ growth in the bottom left panel but normalized to the total median (black curve) across all simulations. This metric shows that, by normalizing the curves by their $M_{\rm star}$ at present, the median LG-like host now had higher fractional mass than the median isolated host *at all redshifts*. The biggest difference occurred at $z \sim 4.2$ (~ 12.3 Gyr ago), where the medians differed by almost an order of magnitude.

While this offset between LG-like and isolated hosts has some host-to-host scatter, we emphasize that at *all* redshifts, the MMP for 4 or 5 of the 6 LG-like hosts was fractionally more massive than the MMP for *all* of the isolated hosts. This result highlights significant systematic differences in the formation histories of the MMPs in LG-like versus isolated environments. This is important when using cosmological zoom-in simulations such as these to compare with observations of the MW and M31 in the LG or with isolated MW-mass hosts as in, for example, the SAGA survey. We also emphasize that the MMPs of LG-like hosts were systematically more massive even back to $z = 7$, indicating these environmental effects manifest themselves early, even when the MMPs were only 10^{-4} to 10^{-3} of their final stellar mass. In Section 2.4.6, we discuss how this likely follows from the mass growth of the DM host halo.

These trends are consistent with a similar analysis of the archaeological star-formation histories (SFHs) of these hosts in Garrison-Kimmel et al. (2019b), which calculated SFHs using all stars in the host galaxy at $z = 0$. By selecting stars in the host galaxy at $z = 0$, and using their formation times alone to calculate the SFH of the host galaxy, this ultimately includes stars that originally could have formed in a galaxy other than the MMP. This is different from how we compute the mass growth curves in Fig. 2.2, where we compute the $M_{\rm star}$ of the MMP at each snapshot. Therefore, this result is not sensitive to the way that one computes the stellar mass growth of the host galaxy.

For comparison, the colored points/bands in Fig. 2.2 (bottom left) show observational inferences of $M_{\rm star}(z)$ in MW/M31-mass galaxies. Using a fit for the SFR main sequence to observational data of star-forming galaxies, Leitner (2012, red circles) constructed mass growth histories, finding that galaxies of $M_{\rm star}(z = 0) \sim 10^{11}$ reached 10 per cent of their final mass at $z \approx 2$ (10.4 Gyr ago),

and 50 per cent of their final mass at $z \sim 1.1$ (8.2 Gyr ago). These values are broadly consistent with our LG-like hosts, though they are earlier than our typical isolated host, but our hosts have somewhat lower $M_{\rm star}(z = 0)$ than in Leitner (2012), so this may not be surprising. In another study, Papovich et al. (2015) used abundance matching to find progenitors of MW/M31-mass galaxies, and calculated their stellar mass evolution, taking into account mass loss from stellar evolution. Galaxies with $M_{\rm star}(z = 0) = 5 \times 10^{10}\,{\rm M}_\odot$ and $M_{\rm star}(z = 0) = 10^{11}\,{\rm M}_\odot$ (akin to the MW and M31) tend to reach 10 per cent of their final mass around $z \sim 2 - 3$ and 50 per cent around $z \sim 1 - 1.5$, respectively. We represent these redshift ranges via the light brown bands in the bottom left panel and note that these results are consistent with the curves for our LG-like hosts. More recently, Hill et al. (2017) also used abundance matching of stellar mass functions at various redshifts to track MW/M31-mass growth histories (e.g. van Dokkum et al., 2013). Hill et al. (2017) found 10 per cent mass occurred at $z \sim 1.3 - 1.9$ (8.9 - 10.2 Gyr ago) for $10^{10.5-11}\,{\rm M}_\odot$ galaxies, and 50 per cent mass occurred at $z \sim 0.6 - 0.9$ (5.8 - 7.4 Gyr ago). We show this range (which brackets our simulation sample) via two dark red bands in Fig. 2.2, which also broadly agree with our simulation suite, though we note that the typical 50 per cent mass in our simulations occurs slightly later than all of these observational estimates. Finally, Behroozi et al. (2019) (pink triangles) used the Bolshoi-Planck DM-only (DMO) simulation, combined with observed properties of galaxies such as stellar mass functions and SFRs to determine the SFHs of galaxies with $M_{\rm 200m} = 10^{12}\,{\rm M}_\odot$ (similar to our sample). They found that these galaxies reached 10 per cent and 50 per cent of their final $M_{\rm star}$ around $z \approx 1.3$ and $z \approx 0.7$ (≈ 8.9 and ≈ 6.4 Gyr ago, respectively), which are consistent with the typical values in our simulations.

2.4.2. What were the building blocks of the galaxy? *Figure 2.3 caption:* **Left:** The number of progenitor galaxies above a given stellar mass versus redshift. Progenitor galaxies are those whose stars end up within the given host-centric distances at $z = 0$: the entire host halo (top), the host galaxy + inner stellar halo (middle), and the inner bulge region (bottom). The lines show the median for our 6 isolated (solid) and 6 LG-like (dashed) hosts, while the shaded regions show the

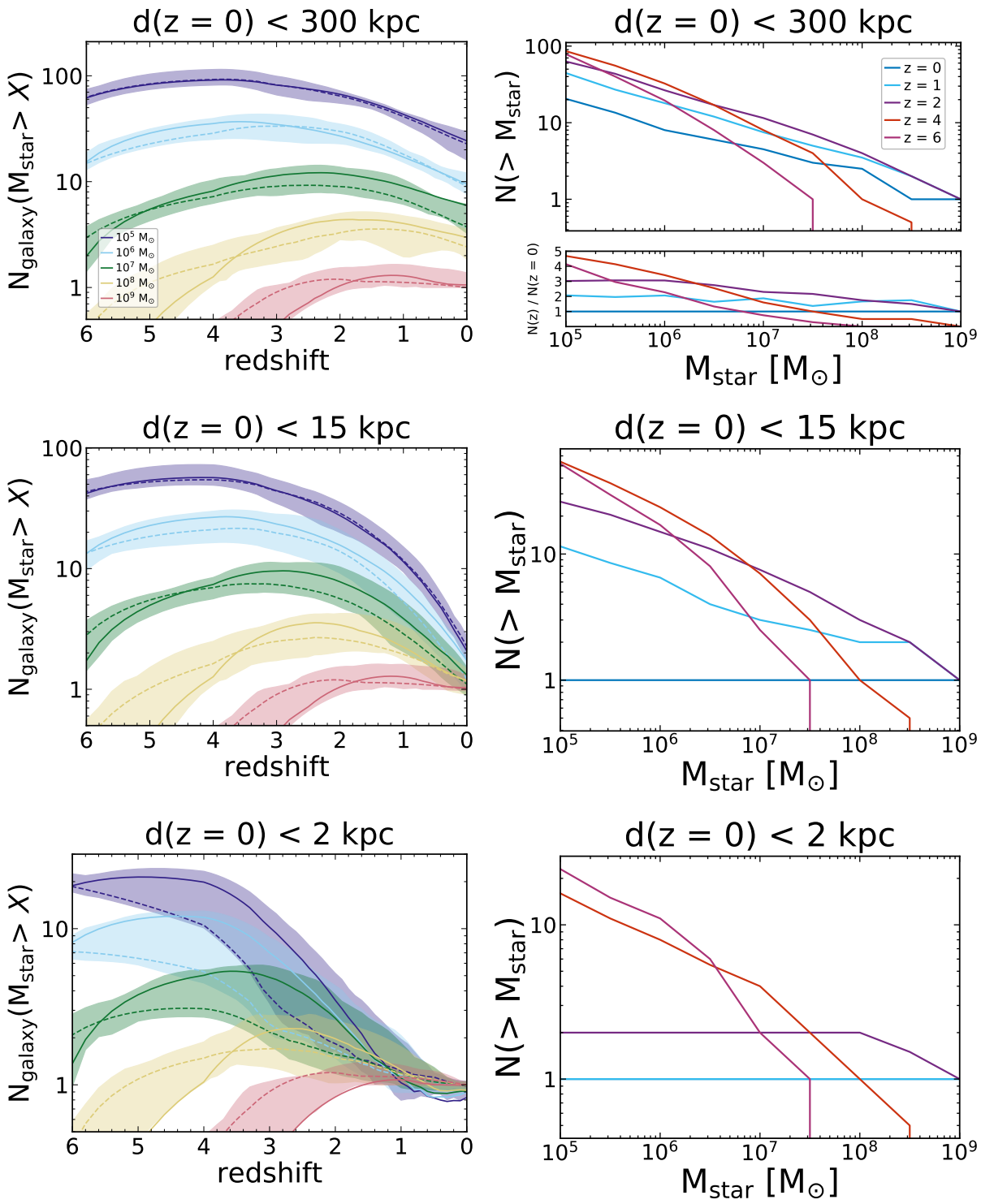

FIGURE 2.3. See the figure caption in the text.

68 per cent scatter across all 12 hosts. Both LG-like and isolated hosts show the same general trends, though the progenitor population of isolated hosts peaks slightly later. **Right:** Cumulative number of progenitor galaxies versus stellar mass at different redshifts, for the median across all 12 hosts. For d_{300} (top), the lower panel shows the number at each redshift normalized to the number at $z = 0$, that is, the population of progenitors relative to surviving satellites at $z = 0$. Note that the progenitor mass function becomes increasingly steeper at higher redshifts (for more discussion see Section 2.5.2). Higher-mass progenitors contributed preferentially at later times. For $M_{\rm star} > 10^5\,{\rm M}_\odot$, $\sim$ 85-100 progenitor galaxies contributed to the formation of a MW/M31-mass system out to its halo virial radius, $\sim$ 55 contributed to the galaxy within 15 kpc, and $\sim$ 20 contributed to the inner bulge region within 2 kpc, with the number of progenitors peaking at $z \sim 4$. Thus, the current satellite population at $M_{\rm star} > 10^5\,{\rm M}_\odot$ represents only $\sim 1/5$ of the population that formed each system.

We next investigate the distribution of 'building blocks' of our MW/M31-mass host galaxies by analyzing the cumulative stellar mass function (number of galaxies above a $M_{\rm star}$ threshold) of all progenitor galaxies (*including* the MMP) across time. We select progenitors that contribute to each of our host-centric distance cuts at $z = 0$. In Fig. 2.3, the left panels show the median number of progenitor galaxies for the isolated (solid lines) and LG-like (dashed lines) hosts versus redshift using $M_{\rm star}$ thresholds. The shaded regions show 68 per cent scatters across our total sample. The panels on the right show the median number of progenitors across the total sample versus $M_{\rm star}$ at given redshifts (we do not show isolated and LG-like hosts separately here).

Fig. 2.3 highlights several interesting trends. For the lowest-mass progenitors that we resolve ($M_{\rm star} > 10^5\,{\rm M}_\odot$), the number of progenitor galaxies peaked at $z \sim 4-5$ (12.2 – 12.6 Gyr ago) for all host distance selections. For d_{300}, the number of progenitors peaked at $\sim 85-100$, while for d_{15} it peaked at $\sim$ 55, and for d_2 it peaked at $\sim$ 20. Given hierarchical structure formation, there were fewer higher-mass progenitors, and their numbers peaked at progressively later times. For example, for d_{300}, the number of progenitors with $M_{\rm star} > 10^7\,{\rm M}_\odot$ peaked at $\sim$ 10 at $z \sim 2.5$. We see

similar trends for 15 and 2 kpc distance cuts, though the progenitor peaks shift to larger redshifts, indicating that the increasingly central regions of the MMP formed earlier. For our highest stellar mass cut, $M_{\rm star} > 10^9\,{\rm M}_\odot$, the hosts typically have only one progenitor (the MMP itself). However, both the MW (with the LMC) and M31 (with M32 and M33) have 1-2 satellites above this mass today. While none of our hosts possess such massive satellites at $z = 0$, half of our hosts (6 of 12) have had at least one satellite with $M_{\rm star} > 10^9\,{\rm M}_\odot$ since $z = 0.7$ (Chapman et al., in prep.), but these massive satellites merge into the host galaxy quickly because of efficient dynamical friction, resulting in the instantaneous median number being 0 across all simulations.

The top right panel also shows the number of progenitors at each redshift normalized to the surviving satellites within 300 kpc at $z = 0$. Compared to the present satellite galaxy population, there were nearly ~ 5 times as many dwarf galaxies that formed each entire MW/M31-mass system, with their numbers peaking at $z \sim 4$ (12.2 Gyr ago). Most of these low-mass progenitors have become disrupted into the main galaxy today. This has implications for galactic archaeology and the "near-far' connection, specifically for any attempt to use present-day nearby dwarf galaxies to make inferences about the high-redshift universe (see Section 2.5.2). Note that the bottom right two panels show just one galaxy (the host galaxy) at $z \sim 0$, because typically these hosts have no resolved satellites inside of 15 kpc (Samuel et al., 2020). For d_{15}, typically 10 progenitor galaxies have contributed to its formation since $z = 1$ (7.8 Gyr ago), while typically only 1-2 progenitors other than the MMP have contributed to the bulge region at $z \lesssim 2$.

We note that for all host distance cuts, the mass function of progenitors was increasingly steeper at higher redshifts which is consistent in both observational and simulation results of the general galaxy population (e.g. Graus et al., 2016; Song et al., 2016; Ma et al., 2018, see Section 2.5.2 for more discussion).

Finally, the satellite populations of isolated and LG-like hosts do not differ substantially across time, as seen in the left panels of Fig. 2.3. For mass bins with $M_{\rm star} \geq 10^7\,{\rm M}_\odot$, LG-like hosts initially had more progenitors, reflective of their earlier formation histories, but are soon outnumbered

compared to the isolated hosts. These differences, however, are small and the right panels show the same behaviour, so we present only the total median in those cases.

Figure 2.4 caption: For all stars that end up within the given host-centric distances at $z = 0$ (top to bottom) and that formed before the given redshifts (different colored lines), panels show the fraction that formed ex-situ (in progenitors other than the MMP) as a function of progenitor stellar mass. The left panels show the (differential) fraction within the $M_{\rm star}$ bin, while the right panels show the cumulative fraction less than $M_{\rm star}$. Lines show total median while shaded regions show 68 per cent scatter across all 12 hosts (see text for discussion of the differences between isolated and LG-like hosts). Because of our sample size, the lower scatter for some mass bins is 0: in these cases we set the lower scatter equal to the median for visual clarity. For all host-centric distances, the ex-situ fraction increases monotonically with increasing redshift and with increasing progenitor mass, so more massive progenitors *always* dominate the ex-situ mass.

Fig. 2.3 shows that the lowest-mass progenitors dominated by number. However, this does not mean that the lowest-mass progenitors dominated the ex-situ mass of the host (that is, the stellar mass that formed in progenitor galaxies other than the MMP). To quantify this, at each redshift we take the population of progenitors in Fig. 2.3 (right panels), excluding the MMP, and we weight each progenitor by their contribution fractions. This gives us the stellar mass in these galaxies, that formed before this redshift, that eventually gets deposited into the host. We then divide this quantity by the total stellar mass in the host at $z = 0$ (for the different distance selections d_{300}, d_{15}, & d_2), to obtain the ex-situ fraction, $f_{\rm ex-situ}(M_{\rm star}, > z)$; we show this in Fig. 2.4. This represents the fraction of stellar mass within the distance selections of the host at $z = 0$, that is above a given age and formed in progenitors of a given $M_{\rm star}$ (other than the MMP). Note that this does not equal the total fraction of $M_{\rm star}$ formed ex-situ *at all redshifts* within each host at $z = 0$ (see Fig. 2.5 for that). Instead, we analyze the ex-situ fraction for stars formed before a given redshift, to highlight trends if one selects stars of minimum age in the MW or M31 today.

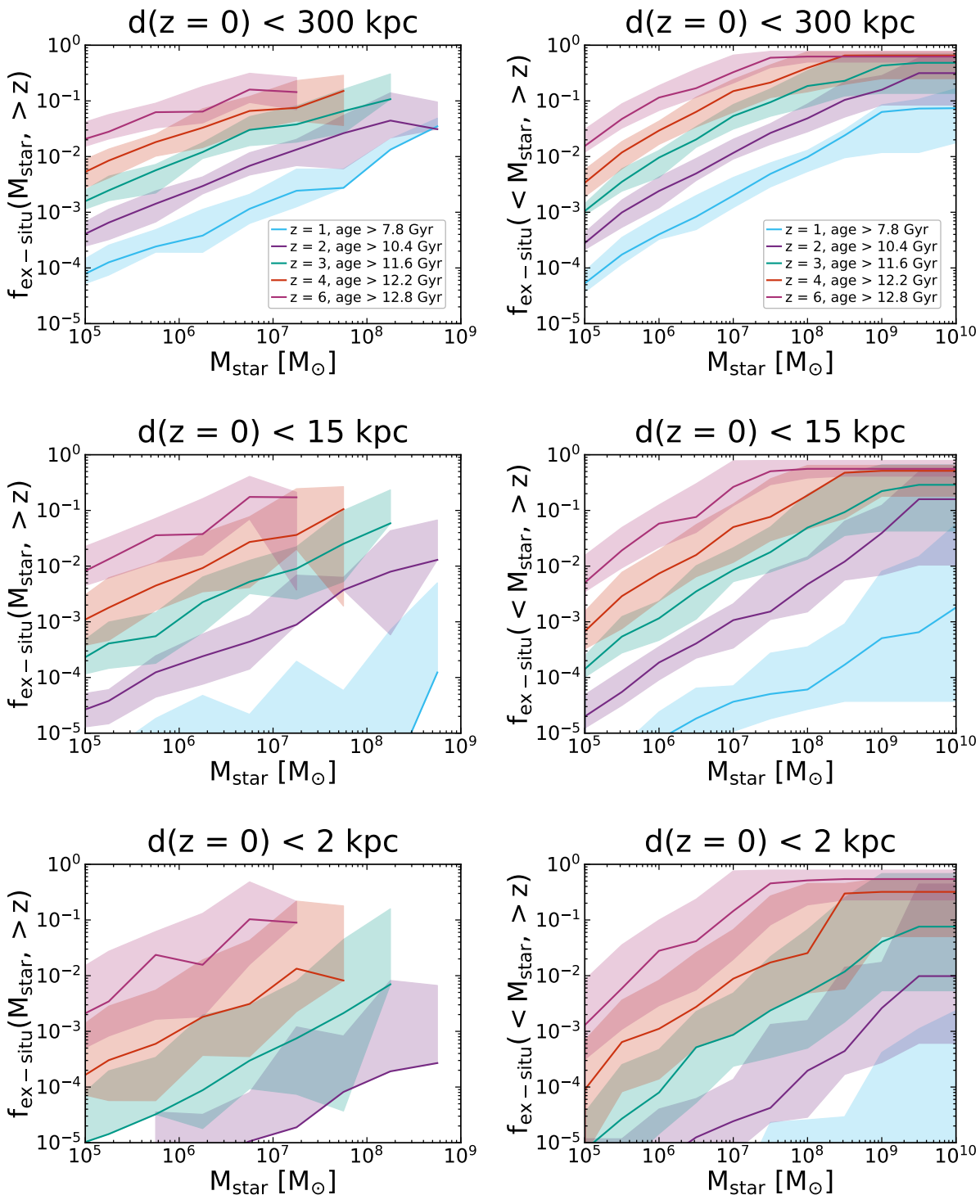

FIGURE 2.4. See the figure caption in the text.

Fig. 2.4 shows $f_{\rm ex-situ}(M_{\rm star}, > z)$ versus progenitor $M_{\rm star}$. The curves in the left panels show the fraction of stellar mass (that formed before a given redshift) that ends up within the three host-centric distances at $z = 0$, that formed inside progenitors other than the MMP, as a function of progenitor stellar mass. The right panels show the cumulative ex-situ fraction that formed in progenitors below a given stellar mass. The shaded regions in all panels show the 68 per cent scatter across the simulations. In some cases on the left panels, given our finite sample of simulations, the lower scatter goes to zero, that is, in some simulations no progenitors at that mass contributed any stars. For visual clarity, we set those values for the lower scatter equal to the median.

Because the slope of the (differential) progenitor mass function, $\mathrm{d}N/\mathrm{d}\log M$, is shallower than unity, weighting progenitors by their $M_{\rm star}$ means that more massive progenitors (other than the MMP) dominated the ex-situ mass, while the contributed $M_{\rm star}$ from the lowest-mass galaxy progenitors was comparatively negligible. Also, as a result of hierarchical structure formation, at later times increasingly more massive progenitors dominated the ex-situ mass. For stellar populations of *all* ages at $z = 0$, the ex-situ mass is dominated by the most massive progenitors, but that mass scale decreases with increasing age.

All of our host-centric distance cuts at $z = 0$ show the same trends above. The primary difference is that regions closer to the center of the host at $z = 0$ have lower ex-situ fractions (normalizations) at all progenitor masses. In other words, the ex-situ contribution becomes increasingly negligible towards the central regions of the host galaxy.

While we do not show it separately in Fig. 2.4 (see instead Fig. 2.5), we find that isolated hosts had larger ex-situ fractions on average than the LG-like hosts. Specifically, the ex-situ fractions for isolated hosts were primarily at the upper end of the 68 per cent scatter region, and LG-like hosts were primarily near the lower end. Thus, ex-situ growth is more important for isolated hosts, but any differences converge by $z \sim 1$.

2.4.3. When did in-situ star formation dominate the mass growth? Having explored the ex-situ fraction as a function of progenitor mass, we next examine the total in-situ fraction for each

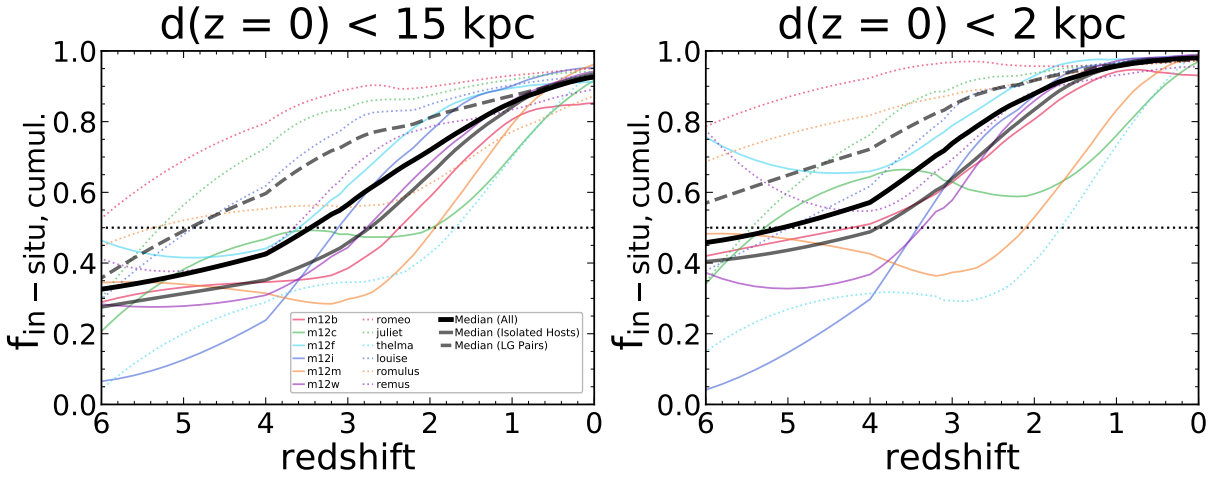

FIGURE 2.5. The cumulative in-situ fraction, defined as the fraction of all stars that formed prior to a given redshift that formed within the most massive progenitor (MMP). Thin colored lines show each simulation, including 6 isolated hosts (solid) and 6 LG-like hosts (dotted). We also show the median across all simulations (thick black line), across isolated hosts (solid grey line), and across LG-like hosts (dashed grey line). The horizontal black dotted line at 0.5 shows our definition for progenitor 'formation', above which the host galaxy has transitioned to having formed the majority of its stars in-situ. Selecting all stars within 15 kpc at $z = 0$, the median progenitor formation across all 12 hosts occurred at $z = 3.5$ (12.0 Gyr ago). For stars in the inner/bulge region, d_2, the main progenitor emerged earlier, at $z = 5.2$ (12.6 Gyr ago). Reflective of Fig. 2.2, for d_{15}, progenitor formation occurred earlier for LG-like hosts ($z = 4.9$, 12.5 Gyr ago) than isolated hosts ($z = 2.7$, 11.3 Gyr ago). For d_2, isolated host progenitors formed at $z = 3.9$ (12.2 Gyr ago), while LG-like hosts were never below 0.5, that is, the majority of stars at all ages in their bulge region today formed in-situ in the MMP. *Thus, main progenitors of LG-like paired hosts formed/emerged significantly earlier.*

host. We aim to understand the transition from (1) an early period when most stars formed ex-situ across several progenitors that (eventually) merge together, to (2) a later period when a single main progenitor dominated the stellar mass growth via in-situ star formation. Thus, we define 'in-situ' star formation as that occurring in the MMP, and we define main progenitor 'formation' (or 'emergence') as the transition from (1) to (2). This is different from other definitions of progenitor 'formation', such as when the galaxy formed a certain percentage of its current $M_{\rm star}$ (see Section 2.4.1). Our goal is instead to quantify when a single main progenitor dominated the stellar mass assembly. To calculate the fraction of in-situ star formation, we select stars within the

host-centric distance selections at $z = 0$ that are older than a given redshift. Of these, we then select the stars that formed inside of the MMP, and divide this stellar mass by the total stellar mass at $z = 0$ that is older than this redshift. Finally, we cumulatively sum these fractions to get the $f_{\rm in-situ,cumul}$, shown in Fig. 2.5.

Fig. 2.5 shows this cumulative in-situ fraction as a function of redshift for the d_{15} and d_2 distance selections. We show each simulation as a thin colored line (solid for isolated hosts, dotted for LG-like hosts), along with medians for the total, isolated, and LG-like hosts. The horizontal dotted line shows a cumulative in-situ fraction of 0.5, which we use to define the formation of a single 'main' progenitor.

Across all 12 hosts, the median formation of the whole galaxy (left panel) occurred at $z \sim 3.5$ (12.0 Gyr ago), and at this redshift, the MMP was ~ 1 per cent of its present $M_{\rm star}$ (see bottom left panel in Fig. 2.2) and the MMP halo was ~ 10 per cent of its present $M_{\rm 200m}$ (Fig. 2.8). Formation in LG-like and isolated hosts occurred around $z = 4.9$ (12.5 Gyr ago) and $z = 2.7$ (11.3 Gyr ago) respectively. However, both host types converged to similar in-situ fractions by $z \lesssim 1$ (7.8 Gyr ago), so these environmental differences were important only in early formation. The median cumulative in-situ fraction at $z = 0$ is ~ 93 per cent, which is consistent with previous results in Anglés-Alcázar et al. (2017), who analyzed the FIRE-1 simulations (including a version of m12i) and found that in-situ star formation dominates ≥ 95 per cent of the stellar mass growth at MW masses. The NIHAO simulations also show similar in-situ fractions (Buck et al., 2019).

Consistent with (radial) inside-out formation, the inner bulge region of the host (right panel) established itself in a single main progenitor earlier than the overall galaxy, with median formation at $z \sim 5.2$ (12.6 Gyr ago). Interestingly, the median for LG-like hosts never extended below 0.5, at least back to $z = 6$ (12.8 Gyr ago), meaning that these stars formed in a single main progenitor *at all redshifts that we probe*. By contrast, bulge stars in isolated hosts formed in a single main progenitor only at $z \lesssim 3.9$ (12.2 Gyr ago).

While we have presented the *cumulative* in-situ fraction, considering all stars that formed to a given redshift, we also examined the *instantaneous* in-situ fraction, using stars that formed within a narrow bin of redshift (not shown). While this is a more time variable/stochastic metric, we found similar trends overall. The key difference is that, because the in-situ fraction rises with decreasing redshift, the cumulative value, being an integral quantity, is smaller than the instantaneous value at a given redshift. Thus, a galaxy transitions above 0.5 at a smaller redshift (typically $\Delta z \sim 0.25$, $\Delta t \sim 0.12$ Gyr) when considering the cumulative in-situ fraction. Both fractions show the same dependence on host-centric distance selection and the same trend that LG-like hosts formed earlier. We also note that the in-situ fraction for the host 'disk' selection, mentioned in Section 2.3.3, is similar to d_{15}, with comparable formation times.

Finally, while we discuss median trends above, we emphasize *significant* host-to-host scatter in Fig. 2.5. For example, for the d_{15} selection, Romeo never had an in-situ fraction below 0.5 back to $z = 6$ (12.8 Gyr ago), while Thelma become dominated by in-situ mass growth only at $z < 1.7$ (9.9 Gyr ago). We also note that the in-situ fractions for some hosts can temporarily decline with time. For instance, in m12c the in-situ fraction decreased from $z \sim 4$ to 2 (12.2 Gyr to 10.4 Gyr ago), which indicates enhanced star formation in progenitors other than the MMP.

Overall, the general trends for both host-centric distance selections are (1) LG-like hosts formed earlier than isolated hosts, (2) the median and scatter tend to converge at $z \lesssim 1$ (7.8 Gyr ago), and (3) no single progenitor forms later than $z \sim 1.7$ (9.9 Gyr ago).

2.4.4. When did a dominant-mass progenitor emerge? Having examined the in-situ fraction of star formation, we also examine a second metric of main progenitor 'formation': when the (instantaneous) stellar mass of the single MMP galaxy dominated that of any other progenitor. Fig. 2.6 shows the ratios of the second- and third-most massive progenitors relative to the MMP, M_2/M_1 and M_3/M_1 (left and right panels respectively), as a function of redshift, for progenitors that contribute stars to d_{15}. We show the median across all 12 hosts (solid purple), 6 isolated hosts (solid green), and 6 LG-like hosts (dashed salmon), as well as their respective 68 per cent scatters

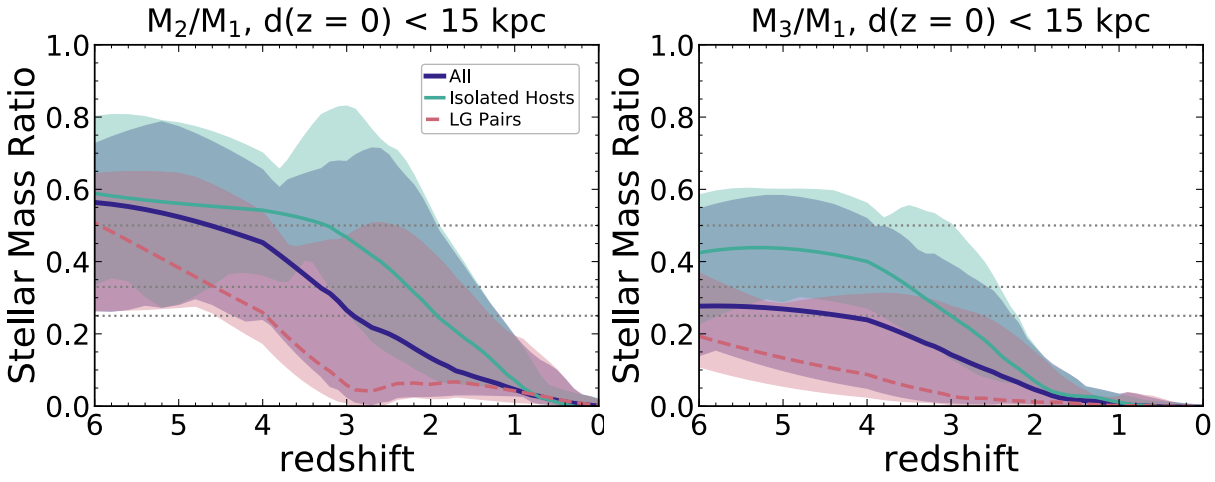

FIGURE 2.6. For all progenitor galaxies that contribute stars to host-centric d_{15}, the ratio of the stellar mass of the second (M_2, left) or third (M_3, right) most massive galaxy to that of the most massive progenitor (MMP, M_1) versus redshift. We show the (smoothed) median across all 12 hosts (solid purple), isolated hosts (solid green), and LG-like hosts (dashed salmon), with the 68 per cent scatter across all hosts in the shaded regions. The horizontal dotted lines show 1:2, 1:3, and 1:4 mass ratios for reference. When each system crosses below these values, the MMP (M_1) increasingly becomes the dominant galaxy across all progenitors. Across all 12 hosts, this transition occurred at $z = 4.6$ (12.5 Gyr ago), $z = 3.3$ (11.8 Gyr ago) and $z = 2.9$ (11.5 Gyr ago) for a 1:2, 1:3, and 1:4 ratio in M_2/M_1. M_3 is rarely within a factor of 2 of M_1, and M_3/M_1 drops below 1:4 at $z = 4.2$ (12.3 Gyr ago). Again, the MMP of LG-like hosts becomes dominant earlier than for isolated hosts: M_2/M_1 for LG-like hosts crossed below 1:2, 1:3, and 1:4 at $z = 6$ (12.8 Gyr ago), $z = 4.6$ (12.5 Gyr ago), and $z = 3.9$ (12.2 Gyr ago), while for isolated hosts these transitions occurred at $z = 3.3$ (11.8 Gyr ago), $z = 2.3$ (10.9 Gyr ago), and $z = 1.9$ (10.2 Gyr ago).

via the shaded regions. Fig. 2.6 also shows stellar mass ratios of 1:2, 1:3, and 1:4 via horizontal dotted black lines.

We find the same qualitative trends in both panels, so we focus primarily on M_2/M_1. However, we note that the median M_3/M_1 ratios were rarely above 1:4 across the total sample, and transitioned below this near $z \sim 4.2$ (12.3 Gyr ago). The (higher) median for the isolated hosts never reached 1:2, and it transitioned below 1:4 later, around $z \sim 3$ (11.6 Gyr ago).

Focusing on the M_2/M_1 ratio (left), we again find that LG-like hosts formed earlier than isolated hosts. Using 1:3 as a fiducial mass ratio, the redshifts where the medians crossed below 1:3 were

z = 4.6 (12.5 Gyr ago) and 2.3 (10.9 Gyr ago) for the LG-like and isolated hosts, respectively, which is consistent with values in the previous section for d_{15}. If we examine all simulations, this transition occurred between the two at z = 3.3 (11.8 Gyr ago). Again, the MMP was $\sim$ 1 per cent of its present $M_{\rm star}$ at these redshifts, meaning that *the main progenitor formed/emerged as the dominant galaxy well before it formed most of its current mass*. For context, the present ratio of $M_{\rm star}$ between the LMC and the MW is about 1:30 ($\sim$ 0.033) and for M33 and M31 it is roughly 1:20, or $\sim$ 0.047 (see compilation in Garrison-Kimmel et al., 2019a).

While we focus on the median trends, we again emphasize the significant scatter in formation histories. We also checked our results using selection of both $d(z = 0) < 300$ and < 2 kpc and saw little change, because the MMP contributes stars throughout the host galaxy at $z = 0$.

2.4.5. When did the main progenitor form? Here, we present the redshifts corresponding to our definitions of progenitor formation regarding the two metrics in Sections 2.4.3 and 2.4.4: the cumulative in-situ star formation fraction and the $\rm M_2/M_1$ (instantaneous) $M_{\rm star}$ ratio. For the in-situ fractions, we use 0.5 as our fiducial threshold to determine when the progenitor formed. Regarding the mass ratio, we choose 1:3 to define formation; this metric does not depend on host-centric distance selection. Fig. 2.7 shows the total median, as well as the medians of isolated and LG-like hosts separately, with the 68 per cent and 95 per cent scatter in the dark and light vertical bars respectively.

Considering main progenitor formation based on in-situ star formation for d_{15}, the median formation for all 12 hosts, isolated hosts, and LG-like hosts occurred at z = 3.5 (12.0 Gyr ago), 2.7 (11.3 Gyr ago), and 4.9 (12.5 Gyr ago) respectively, but the scatters cover a wide range of $z \sim 1.7 - 6$ (9.9 − 12.8 Gyr ago). For d_2, because the median for LG-like hosts never extended below 0.5 at $z < 6$, we set their formation redshift to be $z = 6$ as a lower limit, making the total, isolated, and LG-like host medians z = 5.1 (12.6 Gyr ago), 3.9 (12.2 Gyr ago), and 6 (12.8 Gyr ago), with similar redshift scatter. LG-like hosts had earlier formation times, and we see the same trend of earlier formation for smaller host-centric distance cuts.

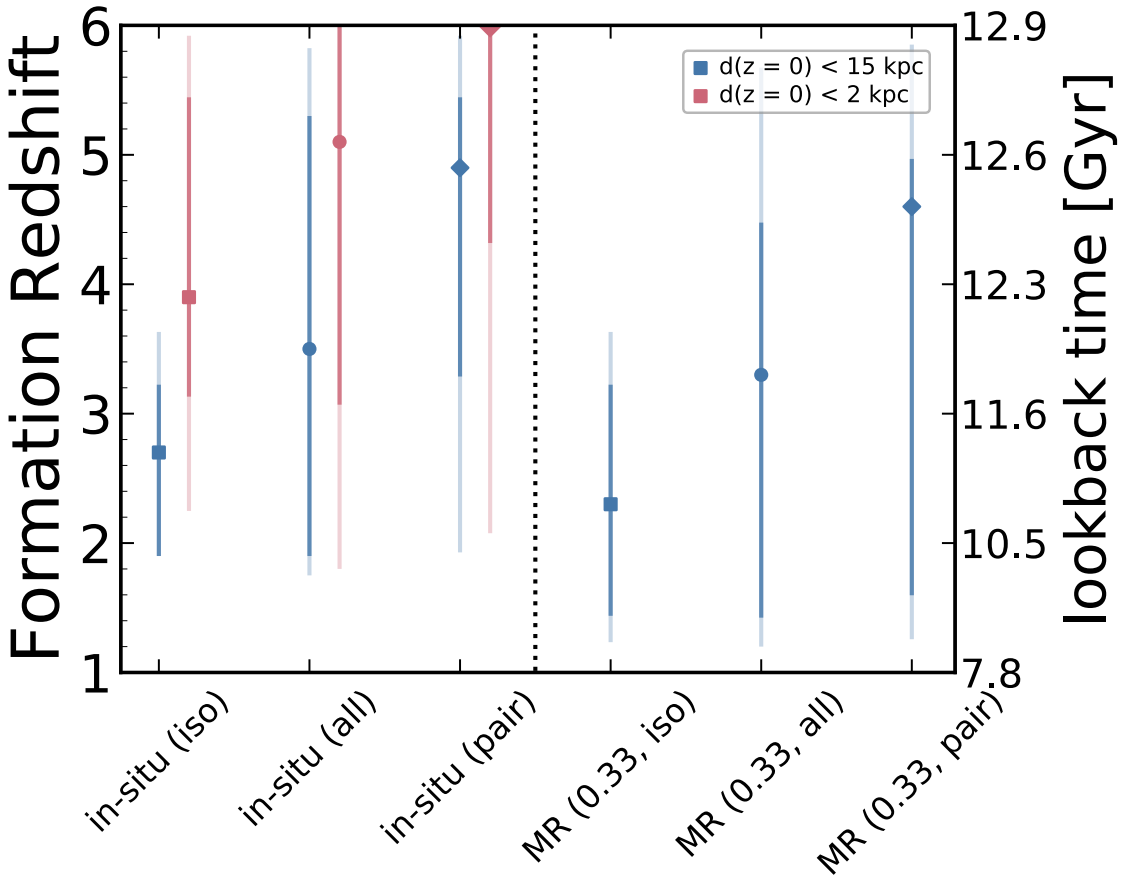

FIGURE 2.7. Summary of the 'formation' redshifts (left axis) or lookback times (right axis) of the main progenitors of our simulated MW/M31-mass galaxies. The left half shows formation defined when the cumulative fraction of stars that formed in-situ exceeds 0.5 (see Section 2.4.3), selecting stars at two host-centric distances at $z = 0$, corresponding to the galaxy + inner stellar halo, d_{15} (blue points), and the inner bulge region, d_2 (red points). The right half shows formation defined when the most massive progenitor exceeds a 3:1 stellar mass ratio with respect to the second most massive progenitor (see Section 2.4.4). Points show the median across the sample and vertical bars show the 68 per cent (darker) and 95 per cent (lighter) scatters, using all 12 hosts (circles), only 6 isolated hosts (squares), and only 6 LG-like hosts (diamonds). For the LG-like hosts at d_2, we show the in-situ formation redshift at $z = 6$ (12.8 Gyr ago) as a lower limit. Considering the entire host galaxy, the main progenitor formation times are similar for in-situ and 3:1 mass-ratio metrics, being $z \sim 3.4$ (11.9 Gyr ago) for the full sample, though significantly later at $z \sim 2.5$ (11.1 Gyr ago) for isolated hosts and earlier at $z \sim 4.7$ (12.5 Gyr ago) for the LG-like paired hosts, with *significant* scatter across different hosts.

If we instead examine main progenitor formation based on instantaneous stellar mass ratio, it reached 1:3 at $z = 3.3$ (11.8 Gyr ago), 2.3 (10.9 Gyr ago), and 4.6 (12.5 Gyr ago) for the total sample, isolated hosts, and LG-like hosts, respectively, with scatter of $z \sim 1.2 - 5.8$ (8.6 – 12.8 Gyr ago). Again, LG-like hosts formed earlier, we find similar median formation times compared with in-situ based formation.

2.4.6. Does the mass growth of the dark-matter halo depend on environment? Finally, we seek to understand more deeply why the most massive progenitor (MMP) of a LG-like host experiences more rapid stellar mass growth (Fig. 2.2) and earlier 'formation' of a main progenitor (Fig. 2.5) than isolated hosts. Specifically, we investigate whether the mass growth of the DM halo reflects these trends as well. We proceed as with Fig. 2.2, but instead we measure the DM halo mass of the MMP galaxy. For most progenitors, the highest-mass galaxy resides in the highest-mass halo. However, because of scatter in stellar versus halo mass growth, especially at early times when there was no clear main progenitor, the MMP galaxy might not reside in the MMP halo. In these cases, we select the halo with the highest DM halo mass as the MMP, but we find nearly identical results if instead we show the DM halo mass of the most massive galaxy.

As Fig. 2.8 shows, we find qualitatively similar results for halo mass growth as for stellar mass growth: LG-like hosts grow in mass more rapidly than isolated hosts. Specifically, across all 12 hosts, the MMP halo reached 10 per cent of its final mass by $z \sim 3.3$ (11.8 Gyr ago), but isolated hosts reached this later at $z \sim 3$ (11.6 Gyr ago) and LG-like hosts reached it earlier at $z \sim 4$ (12.2 Gyr ago), with $\Delta z \sim 1$ and $\Delta t \sim 0.6$ Gyr. At later times ($z \lesssim 2$), we find some enhanced growth for LG-like hosts, but the mass growth histories are more similar, so this environmental effect is weaker for later-term halo growth. These median redshifts are all *earlier* than those for stellar mass in Fig. 2.2, given that central gravitational potential of a DM halo establishes itself earlier than galaxy that it hosts.

We thus conclude that these differences in halo mass growth likely cause (at least to first order) the differences in stellar mass growth and satellite populations, especially because these

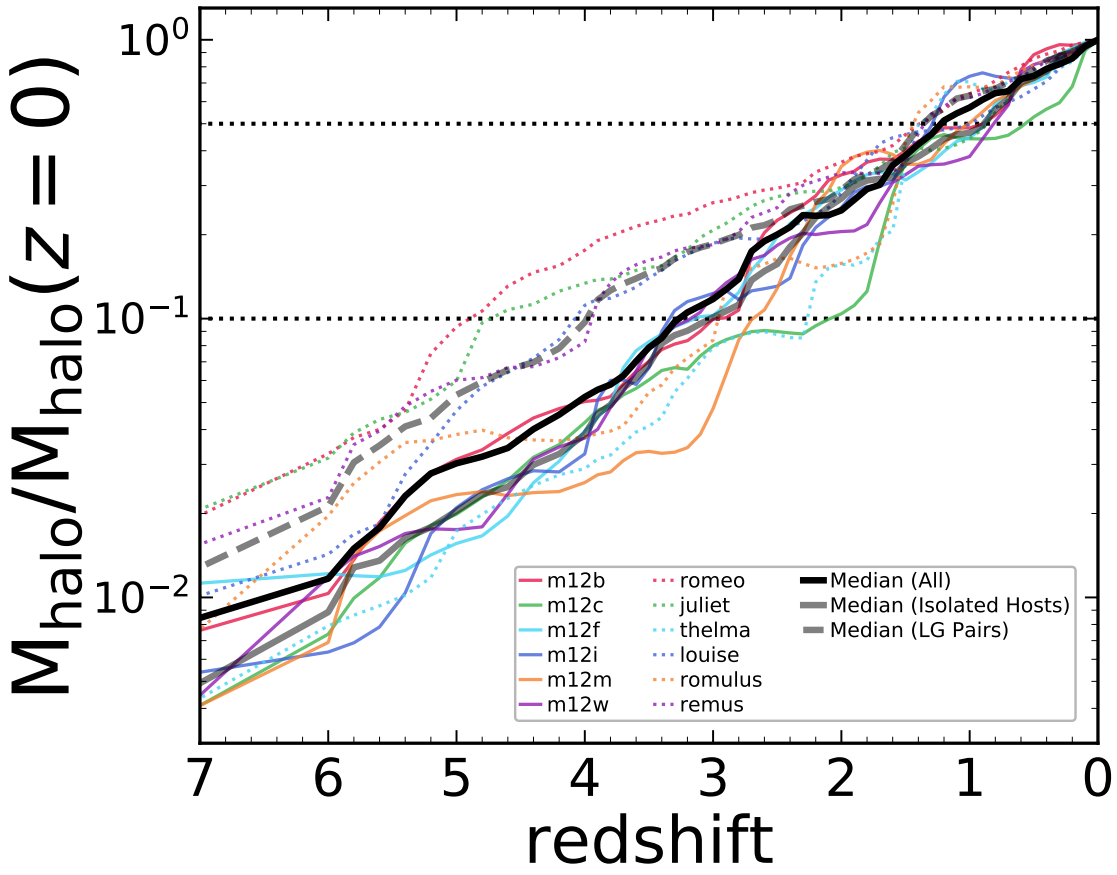

FIGURE 2.8. Similar to Fig. 2.2 (bottom left), but showing the *dark-matter halo* mass of the most massive progenitor (MMP) of each MW/M31-mass host, normalized to each host's M_{200m} at $z = 0$, as a function of redshift (bottom axis) or lookback time (top axis). We show the 6 isolated hosts (solid) and 6 LG-like paired hosts (dotted) in thin colored lines, the median across all 12 hosts (thick solid black), and the medians for isolated (thick solid grey) and LG-like (thick dashed grey) hosts. The dotted horizontal lines show 10 per cent and 50 per cent of the final mass. For the total median, the MMP halo reached 10 per cent of its final mass at $z \sim 3.3$ (11.8 Gyr ago). However, isolated hosts reached this later at $z \sim 3$ (11.6 Gyr ago), while LG-like hosts reached it earlier at $z \sim 4$ (12.2 Gyr ago). At later times ($z \lesssim 2$), the mass growth histories of isolated and LG-like hosts are more similar, so this environmental effect is weaker for later-term growth. We conclude that this more rapid *halo* mass growth in denser proto-LG-like environments at early times likely drives the enhanced *stellar* mass growth in Fig. 2.2.

environmental differences persist back to the initial collapse of theses halos at $z \gtrsim 7$. This is perhaps not surprising, given that halos in LG-like environments formed in denser regions that should collapse earlier than isolated halos (Gallart et al., 2015).

Using a larger sample of 24 paired and 24 isolated host halos in the ELVIS DMO suite of simulations, Garrison-Kimmel et al. (2014) did not find major differences in the median formation times of LG-like halos compared with isolated halos. Similarly, a study by Forero-Romero et al. (2011) using DMO simulations from both the Constrained Local Universe Simulations (CLUES) project and Bolshoi (Riebe et al., 2013), did not see differences in the formation times of isolated versus LG-like hosts. However, both of these works measured halo 'formation' based on the redshift when a halo formed 50 per cent of its final mass at $z = 0$. Garrison-Kimmel et al. (2014) found that both isolated and LG-like hosts had $z_{\rm form} \sim 1.1$, and we find nearly identical results for our baryonic simulations. The key difference in Fig. 2.8 is the *early* formation history of the DM halo, which appears to affect the stellar mass growth to even lower redshifts, as evidenced in Fig. 2.2, where stellar mass growth histories of LG-like versus isolated hosts diverge at $z > 0$.

As mentioned in Section 2.3.1, the simulations assume flat ΛCDM cosmology, and the 6 cosmological parameters span a range of values that are consistent with Planck Collaboration et al. (2018). However, not all simulations used the same cosmology, and one may wonder if this affects their formation times. These differences do not appear to correlate strongly with halo formation time. The Latte suite, Thelma & Louise, and Romulus & Remus adopt the most similar cosmologies. The most distinct cosmology, for Romeo & Juliet and m12w, includes both LG-like hosts and an isolated host, and although Romeo & Juliet did form the earliest (along with Romulus, with $z \sim 6$ for Romeo, $z \sim 5.3$ for Juliet and Romulus), m12w has a relatively late formation time ($z \sim 2.7$). Furthermore, Thelma & Louise span almost the entire range, with Thelma being the latest forming of all hosts and Louise being one of the earliest. Thus we conclude that environment, and not slight differences in cosmology, is the primary cause of the difference in halo/galaxy formation history.

2.5. Summary and Discussion

2.5.1. Summary. Using a suite of 12 FIRE-2 cosmological zoom-in simulations of MW/M31-mass galaxies, we explored their formation histories, to understand when a single main progenitor formed/emerged and quantify the hierarchical build-up from a progenitor population. We defined

main progenitor formation in two ways: (1) when the growth of the MMP transitions from mostly ex-situ to in-situ star formation, and (2) by mass dominance (3:1 ratio or closer) of the MMP compared to other progenitors. The questions that we posed in the introduction and our corresponding answers are:

a) *What were the building blocks (progenitor galaxies) of MW/M31-mass galaxies, and how many were there across cosmic time?*
   i) About 100 progenitor galaxies with $M_{\rm star} \geq 10^5\,{\rm M}_\odot$, ~ 10 with $M_{\rm star} \geq 10^7\,{\rm M}_\odot$, and ~ 1 with $M_{\rm star} \geq 10^9\,{\rm M}_\odot$ formed a typical MW/M31-mass system. Thus, there were ~ 5 times as many dwarf-galaxy progenitors with $M_{\rm star} > 10^5\,{\rm M}_\odot$ at $z \sim 4-6$ ($12.2-12.8$ Gyr ago) than survive to $z = 0$ (Fig. 2.3).
   ii) The slope of the progenitor galaxy mass function was steeper with increasing redshift, which qualitatively agrees with observational and simulation results regarding the overall galaxy population (Fig. 2.3).
   iii) At all redshifts, the ex-situ stellar mass of the accreted population was dominated by the few most massive progenitors, and the ex-situ fraction monotonically increased with redshift (Fig. 2.4).
b) *When did the main progenitor of a MW/M31-mass galaxy form/emerge?*
   i) Across all 12 hosts, a single main progenitor typically formed/emerged around $z \approx 3.3-3.5$ (11.8 - 12.0 Gyr ago) (Figures 2.5, 2.6, 2.7).
   ii) Stars in the inner bulge region formed in a single main progenitor earlier, typically at $z \approx 5.2$ (12.6 Gyr ago) across all 12 hosts (Fig. 2.5).
   iii) Across all 12 hosts, the MMP reached 10 per cent of its present stellar mass by $z = 1.7$ (9.9 Gyr ago) and 50 per cent by $z = 0.5$ (5.1 Gyr ago). Thus, a single main progenitor typically formed/emerged when the host had only a few percent of its final stellar mass (Fig. 2.2).

c) *Does the formation of MW/M31-mass galaxies depend on their environment, specifically, comparing isolated hosts to those in LG-like pairs?*

i) LG-like hosts reached 10 per cent and 50 per cent of their present stellar mass around $z = 2.4$ (11.0 Gyr ago) and $z = 0.8$ (6.9 Gyr ago), respectively. This was significantly earlier than when isolated hosts reached the same fractional masses: $z = 1.5$ (9.4 Gyr ago) for 10 per cent and $z = 0.5$ (5.1 Gyr ago) for 50 per cent (Fig. 2.2).

ii) Similarly, a single main progenitor of a typical LG-like paired host formed significantly earlier ($z_{\rm form} = 4.6 - 4.9$, $\sim$ 12.5 Gyr ago) than for a typical isolated host ($z_{\rm form} = 2.3 - 2.7$, 10.9 - 11.3 Gyr ago) (Figs. 2.5, 2.6, 2.7). This is likely because their DM halos formed earlier (Fig. 2.8).

iii) We find weaker differences between the overall progenitor galaxy populations for LG-like versus isolated hosts across time: the primary difference is that the number of progenitors peaked later for isolated hosts, reflecting their overall later formation histories (Fig. 2.3).

2.5.2. Discussion. Our simulations show that LG-like host galaxies were more massive than isolated hosts (of the same mass at $z \sim 0$), before $z \sim 2$, back to at least $z = 7$. This result is consistent with a related SFH-based analysis of the same simulations in Garrison-Kimmel et al. (2019b). As we showed in Section 2.4.6, this difference is reflected in the early formation histories of the DM host halos, and it may be exacerbated in stellar mass growth if earlier halo formation promotes more metal production throughout the proto-volume, which would make gas cooling more efficient (Garrison-Kimmel et al., 2019b). That paper also found no difference in the SFHs of the satellite galaxies of FIRE-2 isolated versus LG-like hosts, but they *did* find differences in the formation times of central dwarf galaxies in the 'near-field' around LG-like versus isolated hosts, a population that we did not examine in this work.

Our results have key implications for studies of the early Universe and cosmic reionization. The slope of the galaxy luminosity (and mass) function at the faint (low-mass) end informs the

contribution of low-mass galaxies to the ionizing flux during cosmic reionization at $z \gtrsim 7$. For example, if one naively extrapolates the slope to arbitrarily low mass, galaxies with ultra-violet (UV) luminosity $M_{UV} \gtrsim -10$ generated most of the ionizing photons during reionization ($\sim 50-80$ per cent; Weisz & Boylan-Kolchin, 2017). Thus, the evolution of the galaxy luminosity/mass function is of considerable interest. Several observational and theoretical works indicate that the slope of the faint end of the galaxy luminosity/mass function steepens with redshift (e.g. Bouwens et al., 2015; Song et al., 2016). For example, Bouwens et al. (2015) show that the faint-end slope of the UV luminosity function evolves from $\alpha \sim -1.64$ at $z \sim 4$ to $\alpha \sim -2.02$ at $z \sim 8$. Ma et al. (2018) show that in FIRE-2 simulations of larger populations of galaxies, the slope of the low-mass end of the stellar mass function decreases from $\alpha \sim -1.8$ at $z = 6$ to $\alpha \sim -2.13$ at $z = 12$. Graus et al. (2016) summarized several observational works and applied abundance matching to the DMO ELVIS simulations in order to calculate galaxy stellar mass and luminosity functions, finding a slight steepening of faint-end slope from $z = 2$ to $z = 5$.

It is not a priori obvious that the mass function of the progenitors of MW/M31-mass galaxies, which represent a biased region of all galaxies, reflects the overall galaxy population at a given redshift. This is an important question, because the faintest galaxies at $z \sim 7$ will be too faint even for direct JWST observations. Recent works have proposed using resolved stellar populations and SFHs of dwarf galaxies in the LG to infer the faint-end slope of the UV luminosity function at high redshifts, which already provides evidence for a break/rollover in the faint-end slope of the UV luminosity function at $z \sim 7$, given the number of ultra-faint dwarf (UFD) galaxies in the LG (Boylan-Kolchin et al., 2014, 2015; Weisz & Boylan-Kolchin, 2017). On the one hand, our results in Fig. 2.3 (right) show that the progenitors of MW/M31-mass systems do show a similar steeping of mass-function slope to higher redshifts, which is at least qualitatively consistent with the overall galaxy population. However, our results also show that a significant fraction (~ 80 per cent) of these progenitor dwarf galaxies have disrupted into the MW, which means that any inference from the LG population today is missing such progenitors in the early Universe. Thus, our results suggest

that this 'near-far' approach remains promising, but more work is needed to explore quantitatively how representative the proto-LG environments were at high redshifts, which we plan to pursue in future work.

Wide-field surveys currently are measuring elemental abundances, ages, and phase-space distributions for millions of stars throughout all components of the MW, and their sampling rate is expected to continue to increase in the coming decade. A principle aim of these surveys is to reconstruct the formation history of the MW. Our analysis shows that MW/M31-like galaxies were assembled from $\sim$ 100 distinct dwarf galaxies with $M_{\rm star} \geq 10^5\,{\rm M}_\odot$, a majority of which merged by $z \sim 2$ (Fig. 2.3). Each of these dwarf galaxies had a unique orbit, and some likely had distinct elemental abundance patterns. In principle, it may be possible to identify members of many distinct progenitors by identifying them as clumps in a high-dimensional chemo-dynamic space (e.g. Ting et al., 2015), even in the Solar neighborhood (e.g. see recent work by Necib et al., 2019). Achieving this in practice is challenging because progenitors that merged prior to $z \sim 3$ are likely thoroughly phase-mixed by $z = 0$ (see El-Badry et al., 2018b). Such phase-mixing of the earliest-accreted progenitors occurs naturally during merging, and is likely exacerbated by stellar feedback-driven oscillations of the gravitational potential at early times (e.g. El-Badry et al., 2016).

On the other hand, progenitors accreted later ($z \lesssim 2$) – particularly the most massive ones – likely still can be identified. Indeed, there is already compelling evidence that much of the MW's inner stellar halo was formed by a single massive progenitor, whose stars remain dynamically coherent and chemically distinguishable (e.g. Helmi et al., 2018; Belokurov et al., 2018). This fact is not surprising in the context of our simulations: because the progenitor mass functions of MW/M31-mass galaxies are relatively shallow (Fig. 2.3), our simulations generically predict that for any given formation redshift, most of the mass in the stellar halo was contributed by the few most massive progenitors (Fig. 2.4). We finally note that this has also been seen in MW/M31-mass galaxies from the Illustris simulations (e.g. D'Souza & Bell, 2018).

Populations of stars that are both old *and* metal-poor tend to be more centrally concentrated despite the fact that the individual fractions of old *or* metal-poor stars increases with radius from the galactic center (Starkenburg et al., 2017a; El-Badry et al., 2018b). But, because of continued star formation, central regions tend to get crowded over time and the fraction of old/metal-poor stars to the total stellar population becomes vanishingly small. Current stellar surveys (e.g. RAVE, GALAH, APOGEE) thus have a better chance of detecting these stars outside of the solar circle (¿ 8 kpc) and have already found a large number with $[\mathrm{Fe/H}] < -2$. Despite this fact, APOGEE has observed $\sim$ 5100 stars near the Galactic bulge and have found a subset of these stars believed to be both old, but more metal-rich ([Fe/H] ~ -1, Schiavon et al. (2017)). For future work, both within the MW and beyond, LSST also will be capable of finding RR Lyrae stars, which tend to be old (> 10 Gyr), within the LG volume (Oluseyi et al., 2012).

The formation times that we obtain for the entire galaxy (inner stellar halo + disk) are in line with those reported for the halo and thick disk in Gallart et al. (2019) ($z \sim 2 - 4.2$, $\sim 10.5 - 12.3$ Gyr ago). By analyzing the difference in elemental abundances, they determined that the merger between Gaia-Enceladus/Sausage and the MW progenitor was about a 1:4 ratio, and took place roughly 10 Gyr ago, agreeing with previous work (e.g. Helmi et al., 2018; Belokurov et al., 2018; Nogueras-Lara et al., 2019), which would correspond to when the main galaxy was $\sim$ 10 per cent of its total stellar mass, or $\sim$ 30 per cent of its total halo mass. Interestingly, around 10 Gyr the M_2/M_1 ratios in the simulations are more pronounced, at a 1:10 ratio for the total sample, but closer for the isolated hosts, $\sim$ 1:4. Given that the work in Gallart et al. (2019) was done for a LG-like host (the MW itself), assuming Gaia-Enceladus/Sausage was the second most massive component in the system at that time, our results suggest the ratio should be closer to 1:10 or 1:20. At these times, the thick disk was already in place, so this accretion event could have dynamically heated some of these stars into the halo, as well as provide a fresh reservoir of gas for further star formation.

Nogueras-Lara et al. (2019) estimated that the nuclear disk, embedded within the bulge, must have formed over 80 per cent of its stars more than 8 Gyr ago. They also claim that no significant

merger $> 5:1$ occurred in the MW within the last ~ 10 Gyr. Using CMDs of bulge stars to construct SFHs, other studies place the bulge at $\gtrsim 10$ Gyr old, with no traces of a younger stellar population (Zoccali et al., 2003; Valenti et al., 2013; Renzini et al., 2018; Barbuy et al., 2018). Bernard et al. (2018) suggest that 50 per cent of these stars formed before ~ 10 Gyr, and 80 per cent formed before ~ 8 Gyr. The median formation times that we find for the inner bulge region are ~ 2 Gyr earlier than the 10 Gyr lower-limit, however, the scatter does span a wide range from $\sim 10 - 13$ Gyr.

Kruijssen et al. (2019) inferred the MW's assembly history by combining the E-MOSAICS simulation suite, which models globular star cluster formation and evolution in MW/M31-mass galaxies, with the population of observed globular clusters around the MW. They estimate that the MW formed 50 per cent of its $M_{200\mathrm{m}}(z = 0)$ around $z = 1.5$ (9.4 Gyr ago) comparable to both our results and that of Garrison-Kimmel et al. (2014). They also suggest that the main progenitor of the MW must have formed half of its stellar mass by $z \sim 1.2$ (8.6 Gyr ago) and that half of the stellar mass in the main galaxy at $z = 0$ formed *across all progenitors* by $z \sim 1.8$ (10.1 Gyr ago). By selecting globular cluster populations and inferring their evolution through age-metallicity space, Kruijssen et al. (2019) also estimated how many mergers of various masses occurred in the MW. They argue for ~ 15 significant mergers throughout the MW's history, with a majority of them (~ 9) happening before $z = 2$. Although we do not explicitly follow the halo merger tree in our analysis, we see rough consistency in that there were many more progenitor galaxies at high redshift, for all distance selections that we probe.

Although the formation times and mass functions derived in this paper are broadly consistent with other results in the literature, we recognize limitations in our analysis. First, we examine only 12 MW/M31-mass galaxies. A larger sample would allow us to probe better the distributions of formation times across a diverse set of formation histories. We also emphasize again that our galaxies, although comparable to the MW and M31 in many properties, were not created with the intention of exactly reproducing either galaxy. Therefore, one should not interpret the results in

this work as necessarily applying exactly to the MW or M31, but rather as typical cosmological histories for MW/M31-mass galaxies. Finally, our results on stellar mass assembly are contingent on the accuracy and validity of the FIRE-2 model. For example, our assumed meta-galactic UV background from Faucher-Giguère et al. (2009) causes reionization to occur too early ($z \sim 10$) compared to recent observations (e.g. Planck Collaboration et al., 2018, $z \sim 7$); similar results in some other simulation work have reported the same issue (e.g. Oñorbe et al., 2017).

In the future, an interesting follow-up analysis to our results would be a more thorough investigation into the importance of environment on formation time. Although we briefly discussed our initial reasoning of this difference in Section 2.4.6, more work is needed to provide a comprehensive explanation, including a more rigorous comparison of the merger trees between both isolated and LG-like hosts. The dependence of galaxy formation time on halo formation is likely only one piece of this story. Furthermore, there are other important factors that we did not investigate in our analysis, for example, 'patchy' reionization that likely did not happen instantaneously or spatially uniformly, as is implemented in FIRE simulations. Reionization heats gas and suppresses star formation, so a non-uniform reionization would shape a galaxy's SFH differently compared to our simulations (e.g. Lunnan et al., 2012; Zhu et al., 2019). Patchy reionization also could have important consequences for the metal enrichment of gas and early forming Pop III stars. These first stars, and the feedback they produce within small, early forming progenitor galaxies, could affect the SFH of the main host at early times (e.g. Koh & Wise, 2018; Corlies et al., 2018), and a proper implementation of their formation and evolution within simulations is critical. It also remains unclear exactly how Pop III stars enrich the ISM around them, and how they shape the formation of the subsequent generation of stars within the galaxy. Finally, the early formation period of galaxies is governed by many low-mass progenitors (e.g. Ciardi & Ferrara, 2005, also see Fig. 2.1, left column), which are more sensitive to stellar feedback. It remains unclear how dusty these low-mass progenitors are and how that may affect both star formation and gas heating from the UV background. Galaxies that are more dusty can both better shield the reionizing photons

and more effectively cool gas, which causes earlier star formation of lower mass stars than Pop III. Because we have not tested any of these additional points, it is unclear which, if any, would play the dominant role in determining a galaxy's formation time at such high redshifts.

We finally note two other studies focused on galactic archaeology within the FIRE simulation collaboration that are investigating the history and evolution of galaxies like the MW. One example is using chemical tagging of stars to infer where their birth environments were, as well as what other stars they may have co-formed with (Bellardini et al., in prep). Another involves understanding where a population of metal-poor stars currently in the MW disk came from, as noted in Sestito et al. (2019) (Santistevan et al., in prep).

Acknowledgements

The authors thank the anonymous referee for their helpful and thorough report. IBS, AW, and SB received support from NASA, through ATP grant 80NSSC18K1097 and HST grants GO-14734 and AR-15057 from STScI, the Heising-Simons Foundation, and a Hellman Fellowship. JBH acknowledges support from an ARC Laureate Fellowship. MBK acknowledges support from NSF grants AST-1517226, AST-1910346, and CAREER grant AST-1752913 and from NASA grants NNX17AG29G and HST-AR-14282, HST-AR-14554, HST-AR-15006, HST-GO-14191, and HST-GO-15658 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555. CAFG was supported by NSF through grants AST-1517491, AST-1715216, and CAREER award AST-1652522; by NASA through grant 17-ATP17-0067; and by a Cottrell Scholar Award from the Research Corporation for Science Advancement. We ran and analyzed simulations using XSEDE supported by NSF grant ACI-1548562, Blue Waters via allocation PRAC NSF.1713353 supported by the NSF, and NASA HEC Program through the NAS Division at Ames Research Center. Some of this work was performed in part at KITP, supported by NSF grant PHY-1748958.

This work also used various python packages including `numpy` (van der Walt et al., 2011), `scipy` (Virtanen et al., 2020), and `matplotlib` (Hunter, 2007), as well as NASA's Astrophysics Data System.

CHAPTER 3

The origin of metal-poor stars on prograde disk orbits in FIRE simulations of Milky Way-mass galaxies

Published as Isaiah B. Santistevan, Andrew Wetzel, Robyn E. Sanderson, Kareem El-Badry, Jenna Samuel, Claude-André Faucher-Giguère in Monthly Notices of the Royal Astronomical Society, Volume 505, Issue 1, July 2021, Pages 921-938, https://doi.org/10.1093/mnras/stab1345

3.1. Abstract

In hierarchical structure formation, metal-poor stars in and around the Milky Way (MW) originate primarily from mergers of lower-mass galaxies. A common expectation is therefore that metal-poor stars should have isotropic, dispersion-dominated orbits that do not correlate strongly with the MW disk. However, recent observations of stars in the MW show that metal-poor ([Fe/H] $\lesssim -2$) stars are preferentially on prograde orbits with respect to the disk. Using the FIRE-2 suite of cosmological zoom-in simulations of MW/M31-mass galaxies, we investigate the prevalence and origin of prograde metal-poor stars. Almost all (11 of 12) of our simulations have metal-poor stars on preferentially prograde orbits today and throughout most of their history: we thus predict that this is a generic feature of MW/M31-mass galaxies. The typical prograde-to-retrograde ratio is $\sim 2:1$, which depends weakly on stellar metallicity at [Fe/H] $\lesssim -1$. These trends predicted by our simulations agree well with MW observations. Prograde metal-poor stars originate largely from a single LMC/SMC-mass gas-rich merger 7 – 12.5 Gyr ago, which deposited existing metal-poor stars and significant gas on an orbital vector that sparked the formation of

and/or shaped the orientation of a long-lived stellar disk, giving rise to a prograde bias for all low-metallicity stars. We find sub-dominant contributions from in-situ stars formed in the host galaxy before this merger, and in some cases, additional massive mergers. We find few clear correlations between any properties of our MW/M31-mass galaxies at $z = 0$ and the degree of this prograde bias as a result of diverse merger scenarios.

3.2. Introduction

The populations of stars in different regions of a galaxy, such as the disk, the bulge, and the halo, can reveal its formation history. In the standard picture of galaxy formation, unenriched gas collapses in dark matter (DM) halos to form stars and galaxies (e.g. Ostriker & Tremaine, 1975; Rees & Ostriker, 1977; White & Rees, 1978; Fall & Efstathiou, 1980). Many works have studied the hierarchical nature of galaxy formation, such as determining the ages of globular clusters throughout the MW's stellar halo, and the formation of disk galaxies within their DM halos (e.g. Searle & Zinn, 1978; White & Frenk, 1991; Mo et al., 1998). Eggen et al. (1962) first suggested that old, metal-poor stars orbit the MW in elliptical orbits compared to metal-rich stars, which move in more circular orbits, a picture that broadly holds today (e.g. Chiba & Beers, 2000). Many authors have explored how metal-poor stars were deposited into the stellar halo of the MW by mergers of low-mass dwarf galaxies (e.g. Bullock et al., 2001; Bullock & Johnston, 2005; Ibata et al., 1994; Newberg et al., 2003; Nissen & Schuster, 2010). We now understand that most of the mass in the stellar halo came from a few massive mergers, which is a natural consequence of the steepness of the stellar mass halo mass relation. Deason et al. (2016) showed that only $1-2$ satellite galaxies with $M_{\rm star} \sim 10^8 - 10^{10}\,{\rm M}_\odot$ deposited a majority of the stars in the halo, and dwarf galaxies with $M_{\rm star} \sim 10^5 - 10^8\,{\rm M}_\odot$ contributed the bulk of metal-poor ([Fe/H] $\lesssim -2$) stars. A recent observational study by Naidu et al. (2020) suggests that at distances greater than ~ 15 kpc from the galactic center, ~ 80 per cent of halo stars came from 2 mergers: Gaia-Enceladus (Helmi et al., 2018; Belokurov et al., 2018) and Sagittarius (Ibata et al., 1994; Newberg et al., 2003; Majewski et al., 2003).

Many studies focused on the properties and origins of these old and/or metal-poor stars, such as their spatial and elemental distributions within the MW, the masses of the accreted dwarf galaxies they came from, and the formation of the oldest Population III stars (e.g. Scannapieco et al., 2006; Brook et al., 2007; Deason et al., 2016; Griffen et al., 2018). Spectroscopic surveys that have observed stars with [Fe/H] $\lesssim -2$, such as RAVE, GALAH, LAMOST, and APOGEE, provide the elemental abundances and ages of stars in different regions of the MW and now allow for an investigation into the early formation period (Steinmetz et al., 2006; De Silva et al., 2015; Li et al., 2015; Majewski et al., 2017). The Pristine survey also has pushed metallicity measurements down to [Fe/H] < -3, with the potential of reaching the 'ultra metal-poor' ([Fe/H] < -4) population (Starkenburg et al., 2017b). Finally, the recent Hectochelle in the Halo at High Resolution (H3) Survey is targeting stars outside of the disk and bulge regions of the MW, and is reaching metallicities down to [Fe/H] ~ -3 (Conroy et al., 2019).

Recent simulation work on the spatial distributions of old, metal-poor stars within a MW-mass galaxy suggests that they are less centrally concentrated than younger stars. For example, Scannapieco et al. (2006) used DM-only simulations of MW-mass halos and concluded that both unenriched stars, and their descendants, occupy a wide range of radii throughout the MW. In a similar study, Brook et al. (2007) used cosmological baryonic simulations of 4 MW-mass galaxies and found that metal-free stars are distributed throughout the MW's halo, though the authors suggest that the oldest stars are more centrally concentrated. Using a sample of old ($z_{\rm form} > 5$), metal-poor ([Fe/H] < -2) stars in 3 of the FIRE-2 simulations that we examine here, El-Badry et al. (2018b) proposed that this spatially extended distribution of metal-poor stars in a MW-mass galaxy results from a combination of kinematic heating of stars to larger orbits in the host galaxy from a time-varying galactic potential and from early mergers that deposited these stars on radial orbits. However, they did not examine any preference for metal-poor stars to be on prograde orbits aligned with the disk. Similarly, using the APOSTLE simulations, Starkenburg et al. (2017a) found that the fraction of stars with [Fe/H] < -2.5, older than $\gtrsim 13$ Gyr, increases with increasing distance

from the center of the galaxy. Although these studies agree that the number density of old stars is largest near the central bulge region, the overabundance of younger stars can make finding older stars difficult.

The origin of the MW's stellar halo and its disk are likely connected, particularly for the inner stellar halo and the thick disk. Gallart et al. (2019) showed that stellar populations in the inner halo and thick disk partially overlap in kinematic space. The thick disk extends $\gtrsim$ 1 kpc from the midplane and is comprised of primarily older stars across a wide range of iron abundances ($-2 \lesssim$ [Fe/H] $\lesssim -0.5$; e.g. Gilmore & Reid, 1983; Freeman & Bland-Hawthorn, 2002). Several authors have proposed different formation mechanisms for the thick disk such as the kinematic heating of thin disk stars to larger vertical or radial distances, and the disruption and accretion of satellite galaxies (e.g. Walker et al., 1996; Villalobos & Helmi, 2008; Loebman et al., 2011; Haywood et al., 2018) However, many works now show that the disk formed thick and settled over time. Bird et al. (2013) used the cosmological MW-mass simulation 'Eris' (Guedes et al., 2011) and concluded that the MW's disk formed 'upside-down', where older stars formed farther from the midplane with larger vertical velocity dispersions compared to younger stars, which formed in the thin disk with smaller velocity dispersions. More recent cosmological simulations strongly support this idea (e.g. Grand et al., 2016; Bird et al., 2021), including Ma et al. (2017), who analyzed one of the simulated galaxies in our sample (m12i). The authors concluded that older stars in the thick disk formed during the early, bursty phase of the MW when the Galaxy was gas-rich, while younger stars formed during the more calm period of MW growth and that the disk becomes thinner over time.

A great deal of work is dedicated to understanding the existence of net rotation in the stellar halo, and its origins. In an early study, Norris & Ryan (1989) found that, in a population of $\sim$ 500 halo stars, the majority showed net retrograde motion. We now know that the halo is comprised of two components, an inner and outer component, each with different elemental abundances and kinematics (Zinn, 1993; Carney et al., 1996; Bonaca et al., 2017). Deason et al. (2011) found that

metal-rich stars in the halo ([Fe/H] > -2) favored prograde motion and were likely accreted from a satellite, while metal-poor stars ([Fe/H] < -2) were more retrograde on average and could have been part of the early-forming halo. They also noted that many globular clusters in the MW halo orbit in the prograde direction. Another study by Bonaca et al. (2017) concluded that metal-rich stars ([Fe/H] > -1) near the MW have prograde orbits, while metal-poor stars have no average orbit direction. This suggests the metal-rich halo stars formed in-situ in the MW galaxy, while the metal-poor stars likely were accreted. The authors found similar features in one of the simulated galaxies in our sample, m12i. Finally, recent work focused on kinematically, or elementally, coherent structures throughout the MW suggest that many previous mergers currently comprise most of the retrograde population in the MW halo (Myeong et al., 2019; Naidu et al., 2020).

Recently, Sestito et al. (2019) analyzed all known MW stars with [Fe/H] < -4 and concluded that 11 of the 42 stars (26 per cent of their sample) have orbits that are confined to within ± 3 kpc of the disk mid-plane. Of these stars on disk-like orbits, 10 are prograde ($J_\phi > 0$), and only one shows retrograde motion ($J_\phi < 0$). They determined that the remaining 31 metal-poor stars in their sample have halo-like orbits, bringing them far away from the disk, which is typical for accretion-based origin. Subsequently, Sestito et al. (2020) selected stars with [Fe/H] $\lesssim -2$ from both the Pristine survey (Starkenburg et al., 2017b) and LAMOST DR3 (Li et al., 2018), increasing their sample size from 42 to 1027 (a factor of ~ 24). They again found that stars in the prograde region of action space outnumber those in the retrograde region, and ~ 31 per cent of the sample have prograde orbits that are confined to ± 3 kpc of the disk mid-plane. New observational studies continue to support the results from Sestito et al. (2019, 2020), particularly the discovery of metal-poor stars near the plane of the MW (e.g. Venn et al., 2020; Di Matteo et al., 2020). Sestito et al. (2019) posit 3 possible explanations as to the origins of these metal-poor, disk-like stars: (i) they formed in-situ within the MW disk itself and remained on prograde, near-circular orbits; (ii) another galaxy merged into the MW on an orbit that was aligned with the (pre-existing) MW disk, depositing its

stars on prograde orbits; or (iii) these stars formed in one or more progenitor galaxies that merged into the MW as it was forming.

Our goal in this paper is to test the origin and frequency of metal-poor stars on prograde disk orbits, using 12 MW/M31-mass galaxies from the FIRE-2 suite of cosmological zoom-in simulations. The main questions that we address are:

i) How commonly do MW/M31-mass galaxies show a preference for prograde motions in their metal-poor stars, as observed in the MW, and what is the range in strength of this prograde bias?

ii) How does this prograde bias depend on the metallicity, distance, and/or age of the stars?

iii) What process(es) cause this prograde bias among metal-poor stars?

iv) Do any properties of the MW/M31-mass galaxy or their galaxy mergers correlate with this prograde bias?

3.3. Methods

3.3.1. FIRE-2 Simulations.

TABLE 3.1. Properties of the 12 MW/M31-mass galaxies in the FIRE-2 simulation suite that we analyze, ordered by decreasing prograde-to-retrograde ratio for metal-poor stars. Simulations with 'm12' names are isolated hosts from the Latte suite, while the others are from the ELVIS on FIRE suite of LG-like galaxies. We measure all masses and mass fractions at $z = 0$, using only metal-poor ([Fe/H] < -2.5) stars (except for the total stellar mass of the galaxy, which uses all stars). Columns: name; $M_{\rm star,90}$ is the host's stellar mass within $R_{\rm star,90}$, the disk radius enclosing 90 per cent of the stellar mass within 20 kpc; $M_{\rm star,pro}/M_{\rm star,ret}$ is the prograde bias, defined as the mass ratio of prograde to retrograde stars; $t_{\rm lb,merger,1}$ is the lookback time of the merger that contributed the most prograde stars; $f_{\rm merger,1}$ is the fraction of all prograde stars that came from the primary (most contributing) merger; $f_{\rm merger,\ top\ 3}$ is the fraction of prograde stars from the 3 most contributing mergers; $f_{\rm in-situ}$ is the fraction of prograde stars that formed in-situ (within 15 kpc); $f_{\rm merger,1} + f_{\rm in-situ}$ is the sum; $f_{\rm merger,\ top\ 3} + f_{\rm in-situ}$ is the sum of fractions from the 3 most contributing mergers and the fraction that formed in-situ.

Name	$M_{\rm star,90}$ [10^{10} M$_\odot$]	$M_{\rm star,pro}/M_{\rm star,ret}$	$t_{\rm lb,merger,1}$ [Gyr]	$f_{\rm merger,1}$	$f_{\rm merger,\ top\ 3}$	$f_{\rm in-situ}$	$f_{\rm in-situ} + f_{\rm merger,1}$	$f_{\rm merger,\ top\ 3} + f_{\rm in-situ}$
m12m[a]	10.0	9.14	9.56	0.44	0.74	0.21	0.65	0.95
m12w[b]	4.8	3.02	8.17	0.20	0.50	0.13	0.32	0.63
m12c[c]	5.1	2.67	8.99	0.48	0.75	0.21	0.70	0.96
Romeo[c]	5.9	2.35	12.52	0.24	0.38	0.19	0.43	0.57
m12b[c]	7.3	2.29	2.84	0.27	0.59	0.06	0.34	0.65
Juliet[c]	3.3	2.01	12.25	0.14	0.28	0.15	0.30	0.43
Thelma[c]	6.3	1.96	8.61	0.28	0.57	0.10	0.38	0.67
Romulus[d]	8.0	1.56	7.60	0.53	0.77	0.05	0.58	0.82
m12f[e]	6.9	1.48	11.91	0.20	0.35	0.16	0.36	0.51
Remus[d]	4.0	1.42	12.02	0.23	0.45	0.10	0.34	0.55
Louise[c]	2.3	1.28	10.75	0.15	0.34	0.31	0.46	0.65
m12i[f]	5.5	0.98	10.50	0.24	0.57	0.24	0.48	0.81

Note: Simulation introduced in: a: Hopkins et al. (2018), b: Samuel et al. (2020), c: Garrison-Kimmel et al. (2019b), d: Garrison-Kimmel et al. (2019a), e: Garrison-Kimmel et al. (2017b), f: Wetzel et al. (2016).

We use cosmological zoom-in baryonic simulations of MW/M31-mass galaxies from the Feedback In Realistic Environments (FIRE) project[1] (Hopkins et al., 2018). We ran all of the simulations using the same N-body gravitational plus hydrodynamics code Gizmo (Hopkins, 2015), with the mesh-free finite-mass (MFM) hydrodynamics method and the FIRE-2 physics model (Hopkins et al., 2018). Across a temperature range of $10-10^{10}$ K, the FIRE-2 model includes several radiative cooling and heating processes for gas such as free-free emission, photoionization/recombination, Compton scattering, photoelectric, metal-line, molecular, fine-structure, dust-collisional, and cosmic-ray heating, including the spatially uniform, redshift-dependent cosmic UV background from Faucher-Giguère et al. (2009), where HI reionization occurs at $z_{\rm reion} \sim 10$. The simulations self-consistently generate and track 11 elemental abundances (H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe), including the sub-grid diffusion/mixing of these elements in gas via turbulence (Hopkins, 2016; Su et al., 2017; Escala et al., 2018).

Because these FIRE simulations do not model the initial metal enrichment from Pop III stars, we initialize gas elements in the simulations with a metallicity floor of [Fe/H] = −4. We choose this metallicity floor as a rough model for the level of enrichment we expect from Population III stars, which also corresponds to the lowest metallicity in one of the observational studies we compare our results to, Sestito et al. (2019). As we will show, the strength of the prograde bias does not depend on metallicity at low metallicities, so the choice of the metallicity floor is unimportant. Stars form in gas that is self-gravitating, Jeans unstable, molecular (following Krumholz & Gnedin, 2011), and dense ($n_{\rm H}$ ¿ 1000 cm^{-3}). A star particle inherits the mass and elemental abundances from its progenitor gas particle, and represents a single stellar population, assuming a Kroupa (2001) initial mass function. It then evolves along stellar population models from STARBURST99 v7.0 (Leitherer et al., 1999). Furthermore, the FIRE-2 simulations include the following feedback processes: core-collapse and Ia supernovae, stellar winds, radiation pressure, photoionization, and photo-electric heating.

[1]FIRE project web site: `http://fire.northwestern.edu`

We generated the cosmological zoom-in initial conditions for each simulation embedded within periodic cosmological boxes of lengths 70.4 − 172 Mpc using the code MUSIC (Hahn & Abel, 2011), at $z \approx 99$. Each simulation saves of 600 snapshots down to $z = 0$, spaced every ≈ 25 Myr, and all assume flat ΛCDM cosmology with the following range in cosmological parameters, consistent with Planck Collaboration et al. (2018): $\Omega_{\rm m} = 0.266-0.31$, $\Omega_{\rm b} = 0.0449-0.048$, $\sigma_8 = 0.801-0.82$, $n_{\rm s} = 0.961 - 0.97$, and $h = 0.68 - 0.71$.

We analyze the same 12 MW/M31-mass galaxies in Santistevan et al. (2020), and Table 3.1 lists their stellar masses at $z = 0$. Half of our sample comes from the Latte suite of isolated MW/M31-mass galaxies first introduced in Wetzel et al. (2016). The Latte suite has gas and initial star particle masses of 7100 $\rm M_\odot$, but because of stellar mass loss, the average star particle mass is $\approx 5000\,\rm M_\odot$. The mass of DM particles in the zoom-in region is $3.5 \times 10^4\,\rm M_\odot$. The gravitational softening lengths for star and DM particles are fixed at 4 and 40 pc (Plummer equivalent), comoving at $z > 9$ and physical thereafter, while gas elements use adaptive force softening, equal to their hydrodynamic smoothing, which adapts down to 1 pc. We chose the halos in the Latte suite to have $M_{\rm 200m} = 1 - 2 \times 10^{12}\,\rm M_\odot$ and have no other similar-mass halos within $5 \times R_{\rm 200m}$, but we imposed no additional selection criteria beyond this.

The other half of our sample comes from the 'ELVIS on FIRE' suite of LG-like MW+M31 pairs, introduced in Garrison-Kimmel et al. (2019a,b), which have approximately 2× better mass resolution, with gas and initial star particle masses of $3500 - 4000\,\rm M_\odot$. Each pair of halos was chosen based on their masses (each with $M_{\rm 200m} = 1 - 3 \times 10^{12}\,\rm M_\odot$ and total LG mass between $2 - 5 \times 10^{12}\,\rm M_\odot$), current separation (600 − 1000 kpc) and radial velocities at $z = 0$ ($v_{\rm rad} < 0$), and that they have no other massive halos within 2.8 Mpc of either host center.

We selected all halos without prior knowledge of their formation/merger histories or satellite populations, except m12w, which we selected to have an LMC-mass satellite analog at $z \sim 0$ (Samuel et al., 2020). Therefore, our galaxy sample should reflect random/typical formation histories. The host galaxies in our sample reflect a class of morphologies (Garrison-Kimmel et al.,

2018; El-Badry et al., 2018a), which host stellar-to-halo mass relations (Hopkins et al., 2018), disk morphologies and metallicity gradients (Ma et al., 2017; Sanderson et al., 2020), and stellar halos (Bonaca et al., 2017; Sanderson et al., 2018b) akin to that of the MW and M31. Furthermore, each host galaxy possesses realistic satellite galaxy populations, with stellar masses and velocity dispersions (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a), radial and 3-D distributions (Samuel et al., 2020, 2021), and star-formation histories (Garrison-Kimmel et al., 2019b) that broadly agree with observations of the LG and the local Universe.

3.3.2. ROCKSTAR Halo Catalogs & Merger Trees. We generate (sub)halo catalogs using only DM particles, at all 600 snapshots, using the ROCKSTAR 6-D halo finder (Behroozi et al., 2013a). To generate merger trees, we use CONSISTENT-TREES (Behroozi et al., 2013b). All (sub)halos that we examine have zero contamination by low-resolution DM particles, because of the large zoom-in region that we generate for each host.

We assign star particles to (sub)halos in post-processing, which we outline below, but refer the reader to Samuel et al. (2020) for more details. We identify star particles with positions within $0.8 \times R_{\rm halo}$ (out to a maximum distance of 30 kpc) and velocities within $2 \times V_{\rm circ,max}$ of a (sub)halo's center-of-mass velocity. Following these initial criteria, we keep star particles if they are within $1.5 \times R_{\rm star,90}$ (the radius enclosing 90 per cent of the stellar mass) of the current member star particle's center-of-mass and halo center position. This guarantees that the centers-of-mass of both the galaxy and (sub)halo are consistent with one another. We then keep star particles within $2 \times \sigma_{\rm vel,star}$ (the velocity dispersion of current member star particles) of the center-of-mass velocity of member star particles, and we iterate on both criteria until the (sub)halo's stellar mass converges to within 1 per cent.

We use two publicly available analysis packages: HaloAnalysis[2] (Wetzel & Garrison-Kimmel, 2020a) for assigning star particles to halos and for reading and analyzing halo catalogs/trees, and

[2] https://bitbucket.org/awetzel/halo_analysis

GizmoAnalysis[3] (Wetzel & Garrison-Kimmel, 2020b) for reading and analyzing particles from Gizmo snapshots.

3.3.3. Selecting & Tracking Metal-poor Stars. For each MW/M31-mass host galaxy, we define the 'prograde bias' as the ratio of the total stellar mass of prograde stars to that of retrograde stars, $M_{\rm star,pro}/M_{\rm star,ret}$. To measure the prograde bias, we first select star particles in the simulations at $z = 0$ based on the following two criteria, motivated by Sestito et al. (2020): (1) within ± 3 kpc vertically of the disk midplane and within $4 - 12$ kpc radially from the center of the galaxy, which ensures that the sample is not contaminated significantly from bulge stars; (2) with iron abundance [Fe/H] < -2.5. These conditions define our fiducial sample, however, we explore how our results vary with [Fe/H], and with how we spatially and kinematically select star particles in Section 3.4.1. We then define prograde or retrograde motion based on the star particle's action variables, J_ϕ and J_z. J_ϕ is equal to the z-component of angular momentum, L_z. We use the approximation $J_z \approx |z\, v_z|$, where z is a star particle's vertical position and v_z is its vertical velocity, relative to the host's disk. Our choice is motivated by the definition for specific action (per unit mass),

$$J_z \equiv \oint v_z dz \tag{3.1}$$

where the integral is over one orbit in z. In the epicyclic approximation ($J_z \ll |J_\phi|$),

$$J_z \approx \frac{|v_z| z_{\rm max}}{2}. \tag{3.2}$$

Because z follows simple harmonic motion, $\langle |z| \rangle = z_{\rm max}/2$, where $\langle |z| \rangle$ is the time-averaged height. Therefore, for a population of orbits described by the epicyclic approximation, trading time-averaging for spatial-averaging, we get $J_z \approx |v_z\, z|$ for the population. Even though this approximation is good only for stars near the plane of the disk on near-circular orbits, our geometric selection of stars avoids stars that orbit more than 3 kpc above or below the disk at $R = 8$ kpc. As

[3] https://bitbucket.org/awetzel/gizmo_analysis

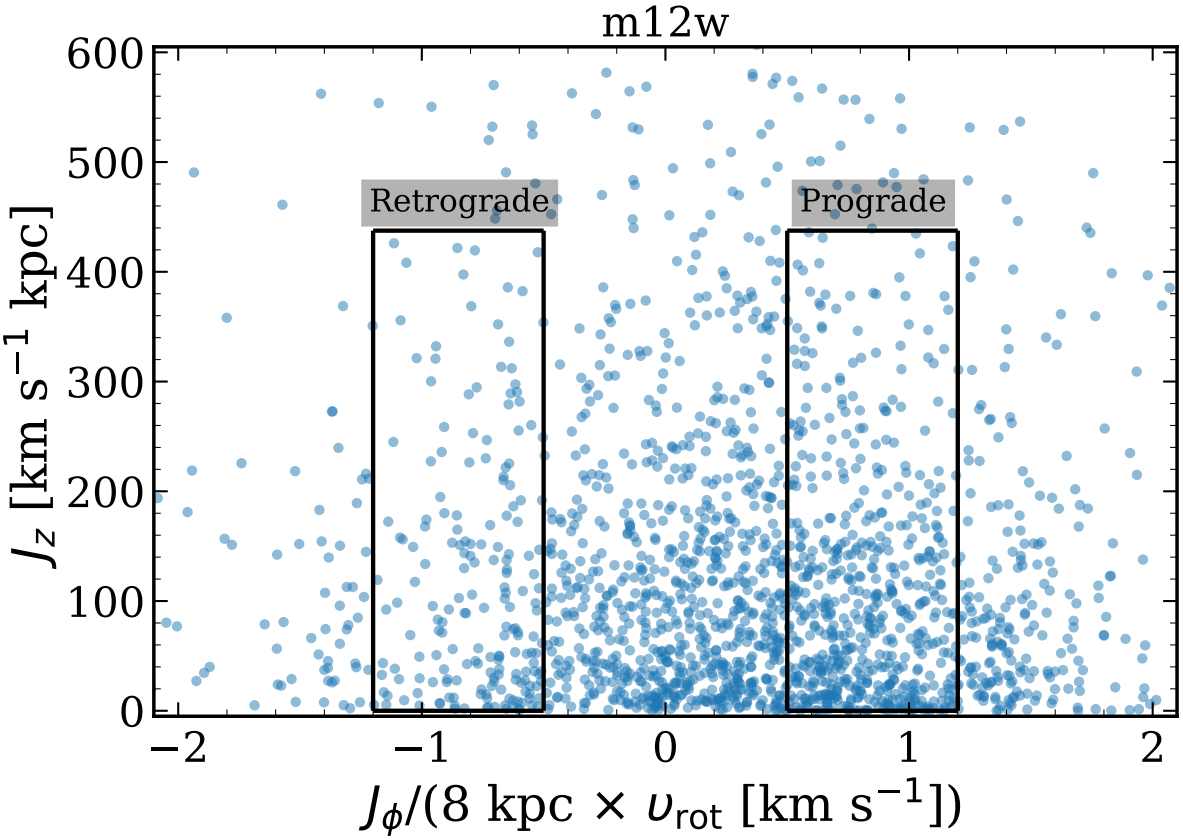

FIGURE 3.1. Metal-poor ([Fe/H] < -2.5) stars in the disk ($4 \leq$ R ≤ 12 kpc, $|$Z$| \leq$ 3 kpc) of the m12w simulation, in action space. Blue dots show individual star particles, and the right and left boxes show regions in action space that we select for prograde and retrograde orbits, respectively. We adopt a similar selection region as in Sestito et al. (2020) for the Pristine survey, which focused on metal-poor stars ([Fe/H] < -2.5) in the MW. We normalize J_ϕ by what 'solar' values would be in the simulations, where $\upsilon_{\rm rot}$ is the rotational velocity of the galaxy at 8 kpc. Metal-poor stars in m12w have a strong preference for prograde orbits, with $M_{\rm star,pro}/M_{\rm star,ret} \approx 3$.

we show in Section 3.1 the strength of the prograde bias does not change much for the different kinematic selections that we investigate.

Sestito et al. (2020) defined prograde stars with $0.5 < J_\phi/J_{\phi\odot} < 1.2$ and $J_z/J_{z\odot} < 1250$, and retrograde stars with $-1.2 < J_\phi/J_{\phi\odot} < -0.5$ and $J_z/J_{z\odot} < 1250$, where the solar values are $J_{\phi\odot} = 2009\,{\rm km\,s^{-1}}$ kpc and $J_{z\odot} = 0.35\,{\rm km\,s^{-1}}$ kpc. Our simulated MW/M31-mass galaxies have a range of rotational velocities ($\upsilon_{\rm rot} \sim 130 - 230\,{\rm km\,s^{-1}}$), so we normalize J_ϕ by $8 \times \upsilon_{\rm rot}\,{\rm km\,s^{-1}}$ kpc,

where $v_{\rm rot}$ is the rotational velocity of the host galaxy, for stars at a similar galactocentric distance as the Sun. Given that we approximate the J_z action variable, we use the same MW values to normalize J_z as in Sestito et al. (2020), and keep stars with $|J_z| < 438\,{\rm km\,s^{-1}}$ kpc. As we will show below, our qualitative results do not depend on the details of our selection window, if for example, we select all prograde ($J_\phi > 0$) and all retrograde ($J_\phi < 0$) stars, or if we vary the J_z cut between $\sim 1/2 - 2\times$ our fiducial value of $|J_z| < 438\,{\rm km\,s^{-1}}$ kpc.

Figure 3.1 shows the action space coordinates of all metal-poor ([Fe/H] < -2.5) star particles within our selection region, for one of our simulated galaxies, m12w. The black rectangles show our fiducial selection windows to define prograde and retrograde stars. An overabundance of metal-poor stars lie in the prograde region (right box) compared to the retrograde region (left box), similar to the MW results in Sestito et al. (2020).

We track these stars, which we select at $z = 0$, back in time to track their origin, specifically, to determine whether they formed (a) in-situ, within 15 kpc physical of the most massive progenitor of the host or (b) in another galaxy that merged into the host. At early times, this distance may encompass other nearby low-mass progenitor galaxies, so we tested how the in-situ fractions changed using stars within 10 and 20 kpc, and we found that they varied by approximately ±4 per cent on average, therefore, our results do not qualitatively change based on this selection. To assign membership to another progenitor galaxy, we require that a star particle must have been a member of that galaxy for at least 3 consecutive snapshots. We also tested requiring 2, 4, or 5 consecutive snapshots, and we found no significant differences in our results.

3.4. Results

3.4.1. How does the prograde bias depend on stellar metallicity, distance, & age?

3.4.1.1. *Comparison with the Pristine survey. Figure 3.2 caption:* The prograde bias, defined as the mass ratio of prograde to retrograde stars, $M_{\rm star,pro}$ / $M_{\rm star,ret}$, versus stellar iron abundance for our 12 MW/M31-mass galaxy simulations at $z = 0$. **Top and middle panels**: the prograde bias as a function of cumulative [Fe/H] and binned [Fe/H], respectively. Solid lines show the 6 isolated

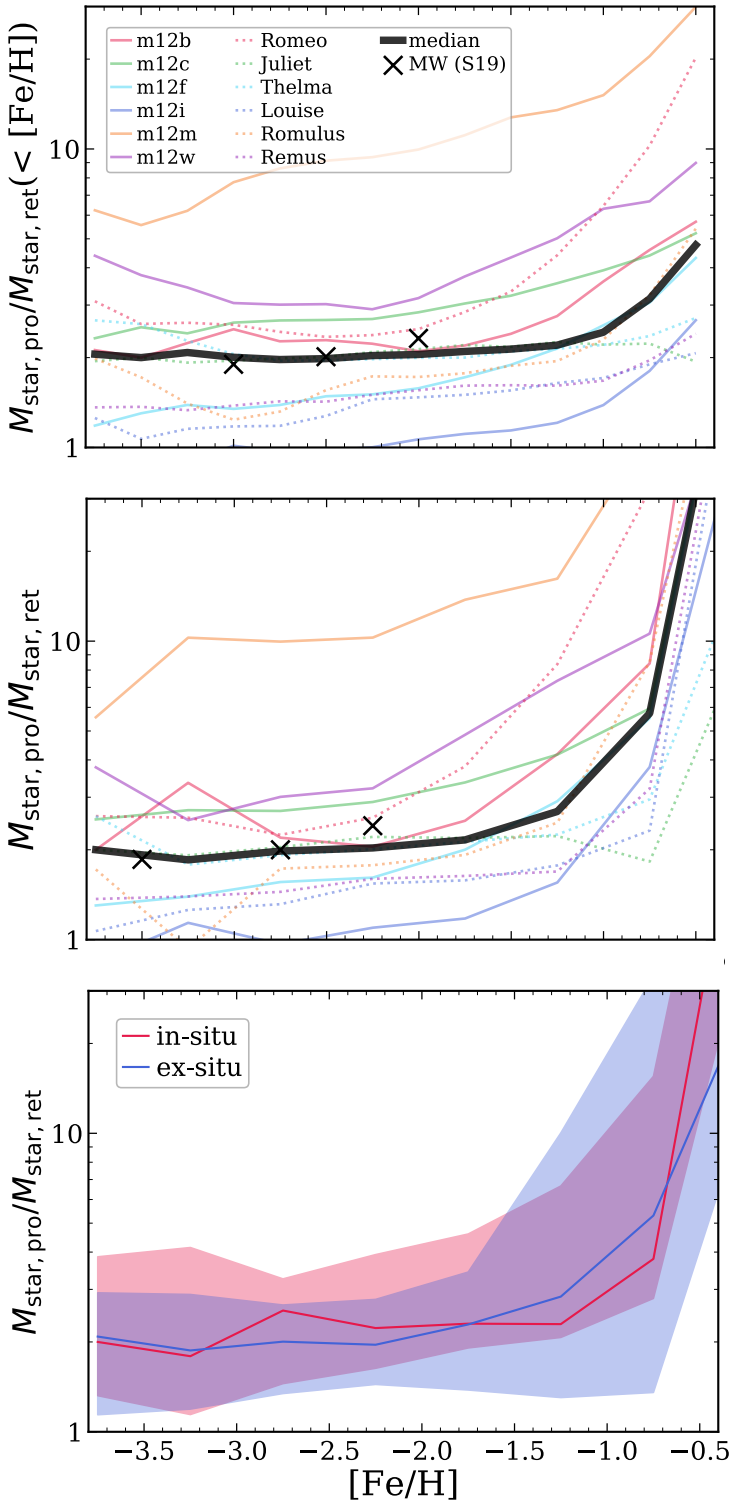

FIGURE 3.2. See the figure caption in the text.

hosts, while dotted lines show the 6 LG-like hosts. The thick black lines show the medians across all 12 hosts, which are consistent with MW observations from the Pristine survey (Sestito et al., 2019, S19), which we show with black crosses. 11 of the 12 galaxies in our sample show a significant prograde bias (> 1) in metal-poor ([Fe/H] < -2.5) stars. For most hosts, the prograde bias is nearly constant with iron abundance at [Fe/H] $\lesssim -1.5$, and it rises rapidly at higher [Fe/H]. **Bottom**: The binned prograde bias, split into in-situ (red) and ex-situ (blue) stars, where we select 'in-situ' stars as formed within 15 kpc of the host. The lines show the medians, and the shaded regions show the 68 per cent scatter across 12 hosts. Ex-situ and in-situ stars have similar prograde bias at [Fe/H] $\lesssim -1.7$, while ex-situ stars show a slightly higher prograde bias at $-1.7 \lesssim$ [Fe/H] $\lesssim -0.7$. As we will show, a single SMC/LMC-mass merger typically drives this trend. At higher [Fe/H], the in-situ component quickly rises in prograde bias, as stars transition to the thin-disk component.

We first examine how the present prograde bias varies with iron abundance, [Fe/H]. Figure 3.2 shows the cumulative prograde bias for all stars below a given [Fe/H], and the prograde bias for stars within 0.5 dex bins of [Fe/H], for all of our hosts, as a function of [Fe/H] at $z = 0$. The thick black lines show the median across all 12 hosts, and we overlay Pristine observations of the MW from Sestito et al. (2019) in black crosses labeled 'MW (S19)'. Solid lines show our 6 isolated hosts, while dashed line show our 6 hosts in LG-like pairs. We show stars only up to [Fe/H] $= -0.5$ to avoid the metal-rich stars in the thin disk that are on overwhelmingly prograde orbits, and we remind the reader of the metallicity floor of [Fe/H] $= -4$ in the simulations.

The median prograde bias is fairly flat at [Fe/H] < -1 for the cumulative plot (top panel) and at [Fe/H] < -1.75 for the binned (middle panel) population. A few hosts (such as m12b) show a slight rise at [Fe/H] $\lesssim -3$. The prograde bias rapidly rises at higher iron abundance, as the stellar population quickly transitions to the more metal-rich, younger, thin disk (and thus highly prograde) component of each host. The observed values from Sestito et al. (2019) are close to our median lines, so the MW is both consistent and typical compared with our simulation suite.

Figure 3.2 shows that nearly all (11 of 12) of our hosts have significant prograde bias at all [Fe/H]. The key exception is m12i, which does not have any significant prograde bias at [Fe/H] $\lesssim -2.5$. However, m12i does have a prograde bias at higher [Fe/H], so this feature is ubiquitous (within our suite) at all [Fe/H] $\gtrsim -2$.

Isolated and LG-like hosts differ slightly in Figure 3.2. Isolated hosts typically have higher prograde bias curves than LG-like hosts, with median values of 2.48 and 1.76 at [Fe/H] = -2.5, respectively. The differences in the prograde biases between isolated and LG-like galaxies is of order ~ 1, which is small compared to the host-to-host scatter, which is of order ~ 10. There are differences in the in-situ formation times between LG-like and isolated galaxies, with LG-like galaxies forming around $z \sim 4.9$ ($t_{\rm lb} \sim 12.5$ Gyr ago) and isolated galaxies forming later, around $z \sim 2.7$ ($t_{\rm lb} \sim 11.3$ Gyr ago), so the lack of strong differences in prograde bias may be surprising (Santistevan et al., 2020). However, as we discuss below, the prograde bias is primarily driven by mergers and sub-dominantly by in-situ formation; we do not explore these differences further throughout the rest of the paper.

Figure 3.2 (bottom panel) shows the same binned prograde bias, but we now split the stars into two origins: in-situ (formed within 15 kpc of the host's main progenitor), and ex-situ stars that formed in another galaxy that merged into the host. The lines show the median values across all 12 hosts, and the shaded regions show the 68 per cent scatter. One might naively expect that the in-situ population would have a larger prograde bias, to the degree that one expects that in-situ stars always formed on disk-like prograde orbits, and because a merger's orbit need not align itself with the host's existing disk. However, we find only weak differences between in-situ and ex-situ stars, which are well within the host-to-host scatter, at all [Fe/H] < -0.5. At [Fe/H] $\lesssim -1.75$, in-situ stars tend to be a little more prograde, but at higher [Fe/H] up to ~ -0.75, ex-situ stars in fact show a slightly stronger preference for prograde orbits. At [Fe/H] > -0.75, in-situ stars quickly become the more prograde population, because we now transition to stars that primarily formed in the host's thin disk.

Figure 3.2 shows that the prograde bias depends only weakly on iron abundance at [Fe/H] $\lesssim -2$. Thus, to maintain consistency with the selection in Sestito et al. (2020), we present results in Section 3.4.2 and after (and Table 3.1) using [Fe/H] < -2.5 as our fiducial selection for metal-poor stars. Our results do not change significantly if we use a slightly different threshold in [Fe/H].

The results in Figure 3.2 do not qualitatively change for different values of the J_z cut as well. When we increase the selection to include stars with $|J_z| < 876\,\mathrm{km\,s^{-1}\,kpc}$, double our fiducial value, the $z = 0$ prograde biases for stars with [Fe/H] < -2.5 change by $\sim 3-6$ per cent. Decreasing the selection by a half to only include stars with $|J_z| < 219\,\mathrm{km\,s^{-1}\,kpc}$ changes the prograde biases by $\sim 3-8$ per cent. Neither of these fractional changes in the prograde bias are significant enough to completely erase the prograde signal, so, we continue to implement our fiducial J_z cut from hereon, unless stated otherwise.

In our calculation of the prograde biases in Table 3.1, the number of prograde star particles range from 341 - 1964, with a mean of ~ 923, and the range of retrograde star particles is 149 - 841, with a mean of ~ 462. These numbers correspond to fractional Poisson uncertainties that range from 2 - 5 per cent for prograde metal-poor star particles, and 3 - 8 per cent for retrograde metal-poor star particles. The mean fractional Poisson uncertainties are 4 and 5 per cent for prograde and retrograde stars, respectively. The typical prograde bias is $\sim 2:1$, therefore, these uncertainties are not strong enough to erase this signal. However, the host with the smallest prograde bias, m12i, is well within these uncertainties of having an equal number of prograde and retrograde stars, and its value of 0.98 should not be interpreted as having a 'retrograde bias'.

Similarly, the Poisson uncertainties for each bin of [Fe/H] are not large enough to significantly change our results. The mean fractional Poisson uncertainty in the number of prograde stars ranges from a min of 0.1 per cent to a max of 9.4 per cent for [Fe/H] $= [-0.5, 0]$ and [Fe/H] $= [-4, -3.5]$ respectively. The uncertainty increases for decreasing [Fe/H] because there are fewer stars in lower metallicity bins. Likewise, the mean fractional Poisson uncertainty in the number of retrograde stars ranges from 0.7 per cent for [Fe/H] $= [-1, -0.5]$ to 13.3 per cent for [Fe/H] $= [-4, -3.5]$.

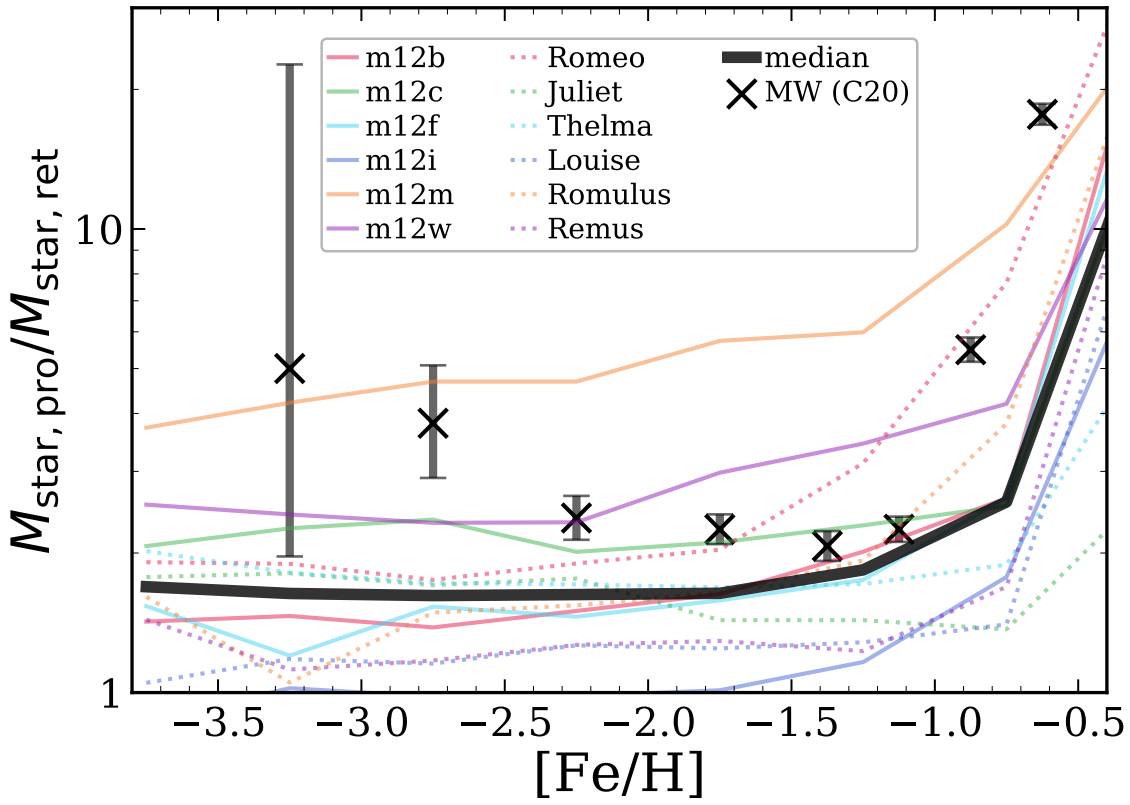

FIGURE 3.3. The prograde bias versus stellar iron abundance, similar to Figure 3.2 (middle panel), but using using a 'stellar halo' selection function to mimic the H3 survey ($1 < |Z| < 3$ kpc and $|b| > 30°$) and computing the prograde bias as the mass ratio of stars with $J_\phi > 0$ to those with $J_\phi < 0$. Crosses show MW observations from the H3 survey (Carter et al., 2021, C20). Despite the differing selection function and prograde metric, the trends in our simulations remain similar to those in Figure 3.2. Our simulations qualitatively agree with the MW observations from C20, though now the prograde bias at [Fe/H] < -2.5 is lower in most simulations than the MW, and most simulations do not show as strong an increase as the MW at [Fe/H] $\gtrsim -1$, although typically at least 2 simulations have prograde biases at least as strong as these MW observations at any iron abundance.

The mean number of retrograde stars in the two highest metallicity bins, [Fe/H] $= [-1, -0.5]$ and [Fe/H] $= [-0.5, 0]$, are similar ($\sim 90,000$), however, the more metal-rich bin has a wider range, which is why it does not have a smaller mean fractional uncertainty.

3.4.1.2. *Comparison with the H3 survey.* Recently, the H3 Spectroscopic Survey (Conroy et al., 2019) also examined the orbits of metal-poor ([Fe/H] < -2) stars in the MW's halo (out of the disk; Carter et al., 2021). They also found a strong prograde bias, with nearly 70 per cent of their stars

on prograde orbits (defined as $J_\phi > 0$), in general agreement with (Sestito et al., 2020), though Carter et al. (2021) found some evidence that the prograde bias increases at lower [Fe/H] as well. Figure 3.3 (crosses with uncertainties) shows their results.

Figure 3.3 also shows the results from our simulation suite when applying a selection window and prograde bias metric similar to Carter et al. (2021). Specifically, we select star particles above the disk, at $1 < |Z| < 3$ kpc, and at high Galactic latitude, $|b| > 30°$. Following Carter et al. (2021), we calculate the prograde bias as simply the ratio of the mass of star particles with $J_\phi > 0$ to those with $J_\phi < 0$. As in Figure 3.2, Figure 3.3 shows each of our hosts separately, and the thick black line shows the median.

The results in Figure 3.3 are similar to those in Figure 3.2 (middle). The median prograde bias is flat across a wider range of iron abundance, $-3.75 <$ [Fe/H] < -0.75, before rising at higher [Fe/H]. The median prograde bias is consistently lower, with a value of ~ 1.6 at [Fe/H] $= -2.5$ compared to ~ 2 in Figure 3.2. By selecting stars at higher vertical distance from the disk and without imposing a selection on J_z, we now select stars with orbits farther from the disk, which leads to a somewhat smaller prograde bias.

Again, our simulation suite broadly agrees with MW observations from Carter et al. (2021), although less well than with the results of Sestito et al. (2020). At $-2.5 <$ [Fe/H] < -1, the observations are consistent with being constant, but they show an upward trend at lower [Fe/H]. The values below [Fe/H] < -2.5 are higher than most of the simulated sample, but the uncertainties encompass much of the region spanned by our simulations. The rise of the MW at [Fe/H] > -1 is sharper than most of the simulations, but the simulations show similar qualitative trends.

To investigate further how this metallicity dependence changes with how we kinematically select star particles, we implemented the H3 kinematic selection ($J_\phi > 0$ versus $J_\phi < 0$) using the Pristine spatial selection window in Section 3.4.1.1 and saw nearly identical results to Figure 3.3. The medians are essentially the same up to [Fe/H] ~ -0.5, where the H3 selection has a slightly smaller value of ~ 10 compared to ~ 15 when using the Pristine window. This suggests that implementing

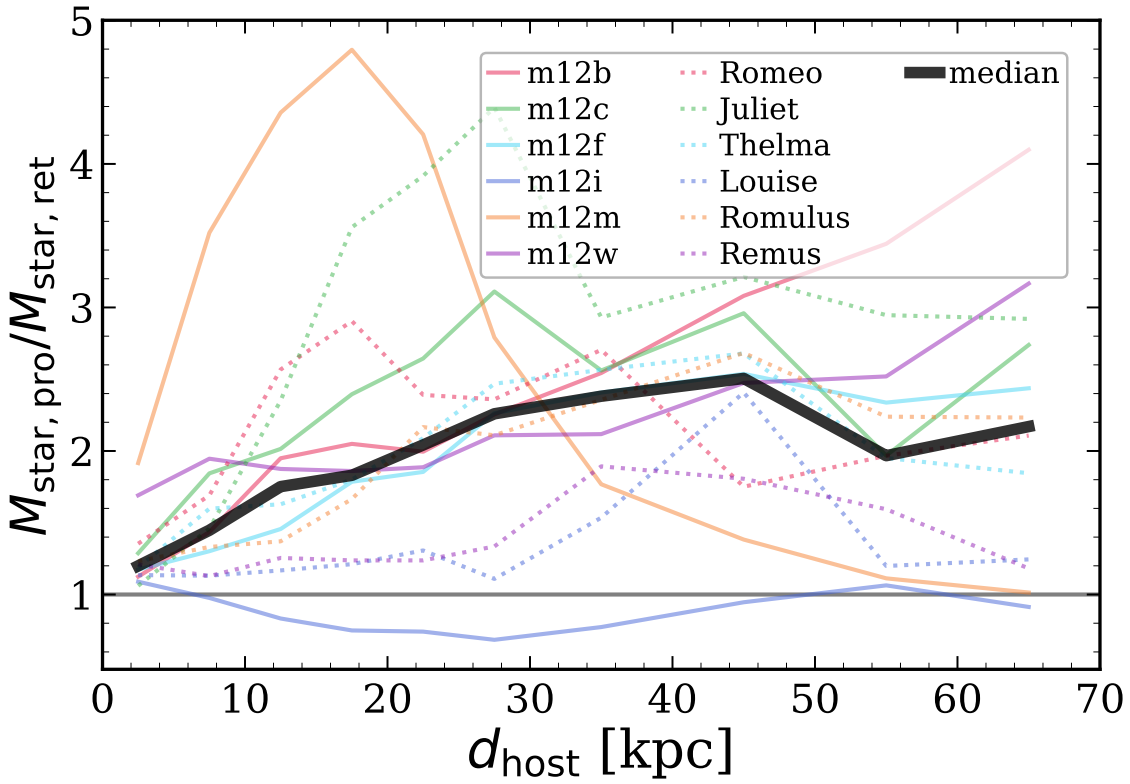

FIGURE 3.4. The prograde bias of the stellar halo. We bin metal-poor ([Fe/H] $<$ -2.5) stars at $z = 0$ based on galacto-centric distance, and we compute the prograde bias as the mass ratio of stars with $J_\phi > 0$ to those with $J_\phi < 0$, relative to the host disk. The black horizontal line at $M_{\rm star,pro}/M_{\rm star,ret} = 1$ denotes no asymmetry, that is, equal mass in prograde and retrograde. Our most extreme hosts, m12m, shows a peak prograde bias of ~ 4.8 at ~ 17.5 kpc before decreasing to 1 at 65 kpc, and m12i again shows no prograde bias now in its stellar halo. Similar to m12m, Juliet shows a peak of ~ 4.5 at ~ 28 kpc, but the prograde bias remains constant at 3, instead of declining. Both Louise and Remus have no appreciable prograde bias until $d \gtrsim 30$ kpc, then they show peaks at $\sim 2-2.5$ between $35-45$ kpc, and no prograde bias at $\gtrsim 65$ kpc. The median prograde bias increases with distance until $d_{\rm host} \sim 45$ kpc, staying roughly constant at ~ 2.5 out to 100 kpc.

a more stringent spatial selection and ignoring the disk midplane, as in the H3 survey, does not change the prograde bias results when compared to our *entire* disk selection. As a final check, we examined how the prograde bias varies using the stellar populations solely above versus below the disk: both populations have nearly identical behavior.

3.4.1.3. *Prograde bias of the stellar halo.* Many studies within the literature debate the possible rotation of the MW's stellar halo, with some arguing for prograde motion and others arguing for retrograde motion. To investigate the possible rotation of the stellar halos in the simulations, we select all metal-poor stars ([Fe/H] < -2.5) and bin them based on their galacto-centric distance, with no additional geometric or kinematic selection. We define the prograde bias in the same way as in the previous section: the stellar mass ratio of stars with $J_\phi > 0$ to those with $J_\phi < 0$.

Figure 3.4 shows how the prograde bias for this 'halo' selection varies as a function of galacto-centric distances, for stars at $z = 0$. The median prograde bias increases with distance to ~ 2.5 at ~ 45 kpc, and the signal flattens at larger distances, up to 100 kpc, though we show only out to 70 kpc. In many hosts, the prograde bias increases with distance up to a certain point, then the prograde bias stays roughly constant. In m12b and m12w, the prograde bias monotonically increases with distance, indicating that stars in the outer regions of these stellar halos are heavily biased towards prograde motion. Our most asymmetric host, m12m, has a peak in its prograde bias at ~ 4.8 for stars at $15 - 20$ kpc, but then the signal decreases and eventually disappears. In our least asymmetric host, m12i, we do not see a prograde bias anywhere, but instead it dips below 1, indicating a retrograde bias at $\sim 5 - 50$ kpc. Finally, Louise and Remus exhibit virtually no prograde bias until ~ 30 kpc, then they increase to $2 - 2.5$ around ~ 40 kpc and fall back to 1 at 65 kpc.

In Figure 3.4, at the solar circle, $d_{\rm host} \sim 8$ kpc, the median prograde bias is ~ 1.5, slightly lower than in Figure 3.3, where the prograde bias is ~ 1.8. Thus, using more stringent selection criteria closer to the disk results in a slightly larger prograde bias. To ensure that this net prograde motion is not driven by stars in the plane of the disk, we also examined a similar trend as Figure 3.4 but only selecting stars at $|Z| > 3$ kpc at all distances, that is, excluding the disk plane. We find nearly similar results as in Figure 3.4. The prograde bias increases until ~ 45 kpc to a value of 2.5, then remains constant. Thus, we find that the details of geometric selection do not influence the results of Figure 3.4 significantly.

Using a similar definition to denote metal-poor stars, [Fe/H] < -2, Deason et al. (2011) suggest that the metal-poor and metal-rich ([Fe/H] > -2) components are rotating in a retrograde and prograde sense, respectively, but the retrograde signal may be a reflection about the adopted value of the local standard of rest. Similarly, Bonaca et al. (2017) suggest that metal-rich halo stars ([Fe/H] > -1) have a net prograde rotation, while metal-poor stars show no net rotation. Although we do not look at a complimentary metal-rich halo population in our analysis ([Fe/H] > -2.5), Figure 3.4 shows a bias for prograde rotation ($\sim 2 - 2.5 : 1$) in our halo selection. Deason et al. (2011) conclude that the metal-rich stars likely were accreted from a satellite, while the metal-poor stars formed in the early host galaxy. This qualitatively agrees with the curves for m12m, Juliet, Louise, and Remus in Figure 3.4, given that the prograde bias peaks at a given distance and subsequently decreases afterward, suggesting possible substructure with net prograde rotation. Bonaca et al. (2017) conclude the opposite, metal-rich stars formed in-situ and metal-poor stars were accreted, which also qualitatively agrees with our results in Section 3.4.2: mergers dominate the origin of the prograde metal-poor stars. Finally, more recent studies, such as Myeong et al. (2019) and Naidu et al. (2020), suggest that the retrograde stellar halo population are a consequence of mergers, though we do not investigate the origins of the retrograde populations in our simulations.

3.4.1.4. *Dependence on stellar age.* In general, a star's metallicity correlates with its age, though with significant scatter. Thus we investigate how the prograde bias at $z = 0$ depends on stellar age, with no selection based on metallicity. We once again select stars as in Section 3.4.1.1, following the approach of Sestito et al. (2020) from the Pristine survey, though we find broadly similar results following the methodology of Carter et al. (2021) from the H3 Survey.

Figure 3.5 shows the the prograde bias at $z = 0$ as a function of stellar age, using age bins 0.5 Gyr wide. The prograde bias increases for younger stars, with a median prograde bias of ~ 4.5 at 10 Gyr of age, as stars transition increasingly to thin-disk formation in each host. This trend is more dramatic for even younger stars (not shown) where the prograde bias reaches ~ 100 for stars of age ~ 7 Gyr. All host galaxies except m12i have a median prograde bias of ~ 2.2 even at the

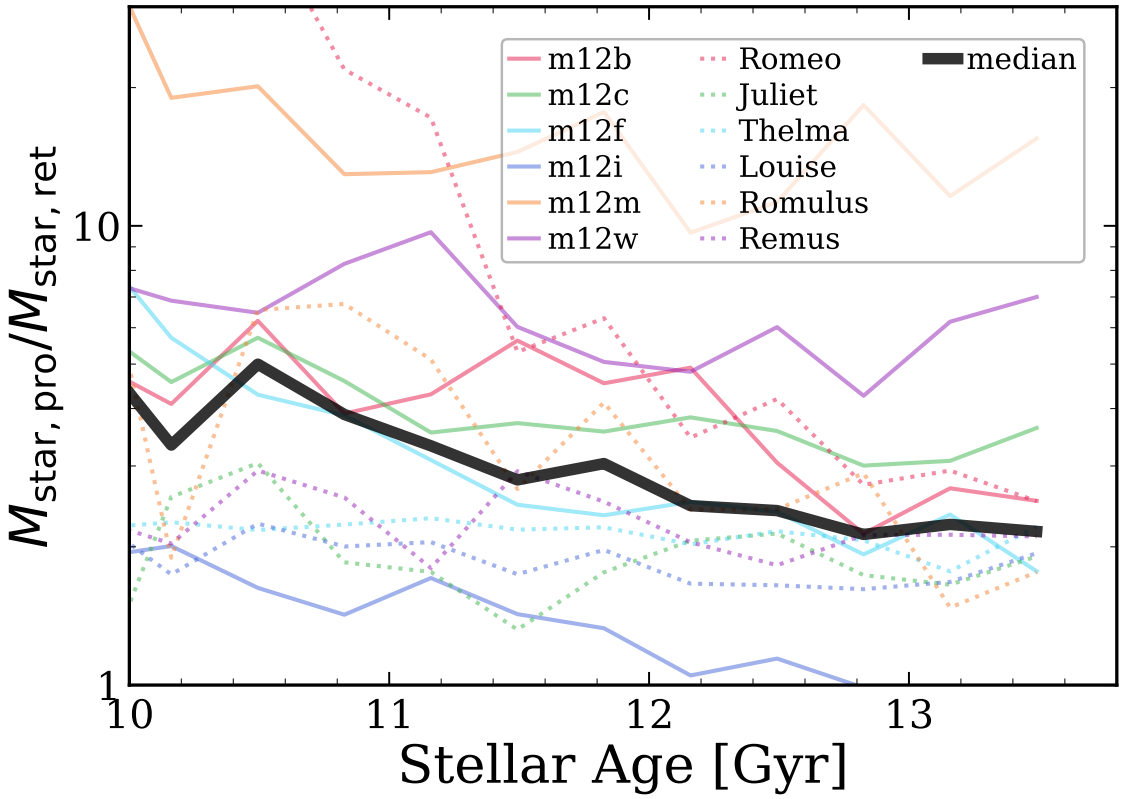

FIGURE 3.5. The prograde bias versus stellar age (regardless of iron abundance) at $z = 0$. Solid lines show the 6 isolated hosts, dotted lines show the 6 LG-like hosts, and the thick black line shows the median across all 12 hosts. The prograde bias decreases with stellar age, such that younger stars tend to be on more prograde orbits, consistent with the expectation that younger stars typically form on diskier orbits. However, even stars in our oldest age bin ($\gtrsim$ 13.5 Gyr) have a median prograde bias of $\sim$ 2.2. All but one host (m12i) shows significant prograde bias to arbitrarily old stellar ages, and all hosts show prograde bias > 1 at stellar ages $\lesssim$ 12.5 Gyr. Thus, the strong preference for prograde orbits in our simulations extends to arbitrarily old and/or arbitrarily metal-poor stars.

oldest ages ($\gtrsim$ 13.5 Gyr), and m12i does show prograde bias starting at $\sim$ 12.5 Gyr age. We thus conclude that the prograde bias at $z = 0$ is nearly ubiquitous for star at all ages, including arbitrarily old stars. While this result might be surprising, as we will show below, it can be understood as arising from the fact that the prograde bias arises largely from a single major merger, which brings in old stars with it.

The oldest stars ($t_{\rm form} > 13$ Gyr) have a wide range in iron abundance, [Fe/H] = $[-3.87, -0.52]$, with a median value of [Fe/H] = −2.5. This is broadly consistent with El-Badry et al. (2018b), who used 3 of the simulations in our sample and found that stars that form at $z_{\rm form} > 5$ ($t_{\rm lb} > 12.6$ Gyr) have a range of [Fe/H] = $[-3, -1]$, while stars with [Fe/H] ~ -2 formed at $1.5 \lesssim z \lesssim 8$ (9.4 Gyr $\lesssim t_{\rm lb} \lesssim$ 13.1 Gyr).

3.4.2. Which galaxies contributed to the prograde bias? We next investigate the origin of the prograde bias in metal-poor stars, where we define 'metal-poor' as [Fe/H] < −2.5 hereon. Using the tracking method in Section 3.3.3, we first attribute all star particles as either formed in-situ (formation within 15 kpc from the host) or ex-situ. For ex-situ star particles, we track which progenitor galaxy they were a member of prior to merging into the host, and we track the total stellar mass that each progenitor contributed to the prograde metal-poor population of each host. We examine the three progenitors that contributed the most mass to the prograde metal-poor stars (or, put differently, the three progenitors that contributed the largest number of prograde metal-poor stars), and we refer to the galaxy that contributed the most, second most, and third most as the 'primary', 'secondary', and 'tertiary' merger, respectively. These mergers, however, are not always the top three most massive mergers to have ever contributed prograde metal-poor stars to the host. Table 3.1 lists values for the in-situ fractions, $f_{\rm in-situ}$, primary merger fractions, $f_{\rm merger,1}$, and fraction of stars contributed from the primary, secondary, and tertiary mergers combined, $f_{\rm merger,top3}$. Perhaps surprisingly, none of these fractions show any clear correlation with the strength of the prograde bias.

For all hosts except Louise and Juliet, the primary merger fractions are higher than the in-situ fractions, with median values of $f_{\rm merger,1} \approx 0.24$ and $f_{\rm in-situ} \approx 0.16$. *Thus, the primary merger generally was responsible for more prograde metal-poor stars than the (most massive progenitor of the) host galaxy.* The primary merger fractions range from $\approx 15 - 55$ per cent: the merger in Romulus contributed the most, and the mergers in Juliet and Louise contributed the least. The in-situ fractions range from $\approx 5 - 30$ per cent. Furthermore, on average, the primary merger

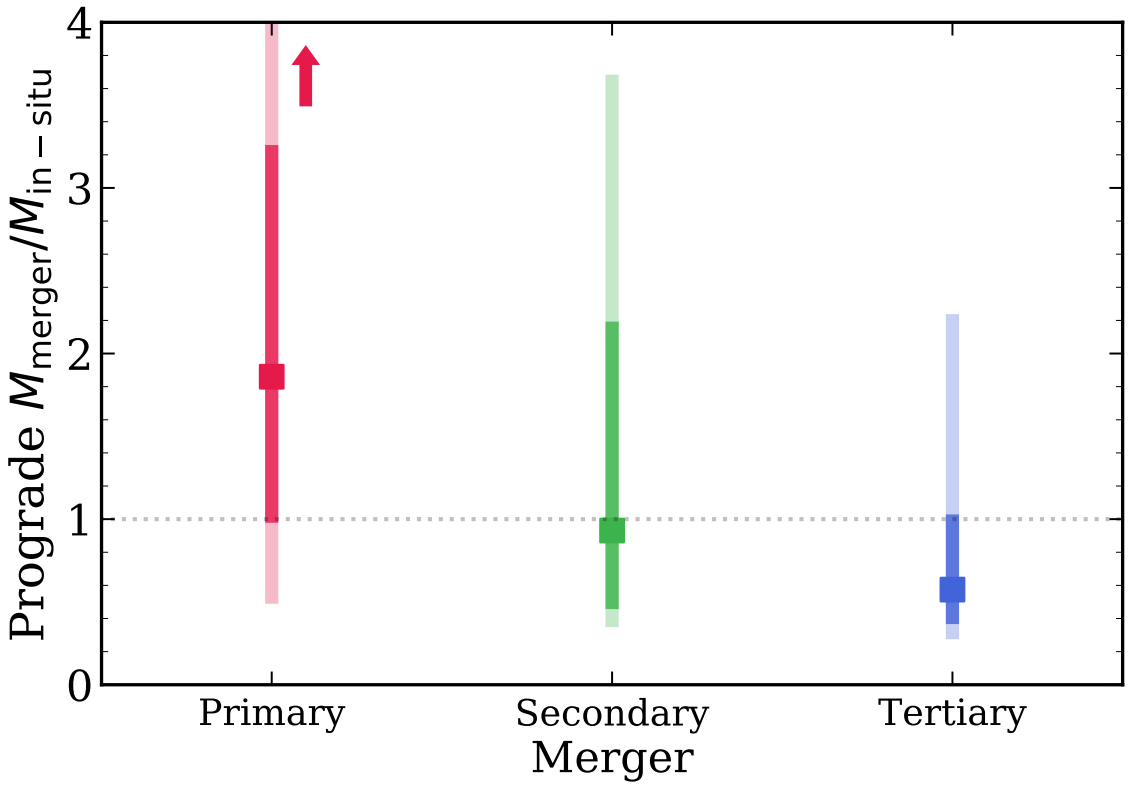

FIGURE 3.6. The ratio of the mass of prograde metal-poor stars contributed from a merger to that from in-situ formation, for the primary (red), secondary (green), and tertiary (blue) galaxy mergers, defined as the 3 progenitor galaxies that deposited the most mass to the prograde metal-poor population. Squares show the median values across all 12 hosts, and the dark and light vertical bars indicate the 68 per cent scatter and full range, respectively. The red arrow indicates that the full range of values for the primary-to-in-situ ratios extend beyond the axis (to $\sim$ 11). Relative to in-situ stars, the primary merger typically contributed $\approx 2\times$ more stars, the secondary merger was comparable to in-situ, and the tertiary merger contributed only $\sim 1/2$ as much as in-situ. The sum of these 4 components nearly always accounts for the majority of prograde metal-poor stars (see also Table 3.1), and any single merger beyond the tertiary merger is typically negligible. Thus, the primary source of prograde metal-poor stars at $z = 0$ is from a single galaxy merger.

contributed more prograde stars ($\sim$ 24 per cent) than retrograde stars ($\sim$ 16 per cent) to the host galaxy.

Figure 3.6 shows the ratio of the stellar masses contributed by the primary (red), secondary (green), and tertiary (blue) mergers, relative to in-situ formation, of prograde metal-poor stars at $z = 0$. The square points show the median values across our 12 hosts, and the dark and light

vertical bands show the 68 per cent scatter and full range, respectively. The full scatter extends up to ≈ 10 in the primary merger to in-situ ratio, though we truncate the plot for clarity. The primary merger contributed twice as many stars as in-situ formation, so *the primary merger is the single most important contributor to prograde metal-poor stars*. Furthermore, the secondary merger contributed comparable stellar mass as in-situ formation. Finally, the tertiary merger contributed typically only $\sim 1/2$ that of in-situ formation. We emphasize that most metal-poor stars formed early, $\gtrsim 12\,\mathrm{Gyr}$ ago; this includes both stars that form in-situ and ex-situ. Almost all metal-poor stars formed prior to these 3 mergers ($\gtrsim 93$ per cent), thus, these mergers did not induce in-situ formation that contributed to the prograde bias.

While Figure 3.6 shows the masses of the mergers and in-situ stars that comprise the prograde population, we also investigated the total stellar masses of these mergers relative to each other. Specifically, we computed the ratio of the peak stellar masses (across their histories) between the primary to the secondary, and the secondary to the tertiary. We found that the peak stellar mass ratio between the secondary to the tertiary was ~ 2, consistent with Figure 3.6, but the ratio between the primary to the secondary was ~ 6, greater than in Figure 3.6. This implies that the primary merger contributes a smaller *fraction* of its original stars to the prograde metal-poor population compared to the secondary merger, even though it contributes more prograde stars in total.

Table 3.1 shows that in-situ formation together with the primary merger account for $30-70$ per cent of all prograde metal-poor stars. Considering the in-situ fraction along with the top 3 mergers, this accounts for $43-96$ per cent. Juliet has the smallest summed fraction, so an unusually large number of additional mergers contributed significantly to its prograde metal-poor stars. In m12m, m12c, and Romulus, the combined fractions reach ≈ 82 per cent or higher, and the top 3 mergers account for roughly ≈ 75 per cent. We also find no clear differences in these fractions for isolated versus paired hosts; m12c (isolated) and Romulus (LG-like) have the 2 highest primary merger fractions, and m12b (isolated) and the Romulus have the 2 lowest in-situ fractions. Furthermore, m12m, which has the highest prograde bias, has a relatively high primary merger fraction and

average in-situ fraction, while m12i, which has the smallest (no) prograde bias, has average values for both. These results highlight the diversity of merger/growth histories that lead to a prograde bias for metal poor stars at $z = 0$, and that the fractions of such stars from in-situ formation or the primary merger do not predict the strength of this kinematic feature.

We thus conclude that mergers dominate the origin of prograde metal-poor stars, and the primary merger dominates over any other merger. We thus focus most of our subsequent analysis on the properties of the primary merger.

3.4.3. How did the prograde bias evolve over time? *Figure 3.7 caption:* The formation histories for 6 galaxies in our sample. We show half of our simulated galaxies, which represent the diverse origins of the prograde bias across our suite; the other galaxies have similar behavior. Arrows indicate the times of the top three galaxy mergers, ranked by the mass that they contribute to prograde metal-poor ([Fe/H] < -2.5) stars at $z = 0$, with arrow darkness indicating the rank, such that the primary merger is the darkest. For each galaxy, the upper sub-panel shows the evolution of the prograde bias: the dotted horizontal line at 1 indicates no bias (equal mass in prograde and retrograde stars). At early times, the prograde bias was typically near 1, though this reflects the fact that the disk had not formed yet, so prograde and retrograde are not well defined; see discussion below. For all galaxies except m12i (top right), the prograde bias increased sharply following either the primary merger (as in m12m, m12w, m12b, Romulus) or the secondary merger (as in Romeo). m12w also shows an additional increase after the tertiary merger. For each galaxy, the lower sub-panel shows the evolution of the offset angle between the angular momentum vector of the galaxy's disk and the angular momentum vector of the primary merger's orbit, where we freeze the latter at its final orientation vector after the merger coalesces with the host (see Section 3.4.4). At early times, the offset angle of the host's disk was un-aligned with the (future) merger, and it rapidly shifted given the high merger/accretion rates, and that the disk itself was only marginally defined. However, the host's disk typically aligned itself with the orbital plane of the primary merger (as in m12m, m12w, m12b, Romulus) during/after the merger event. These mergers were typically

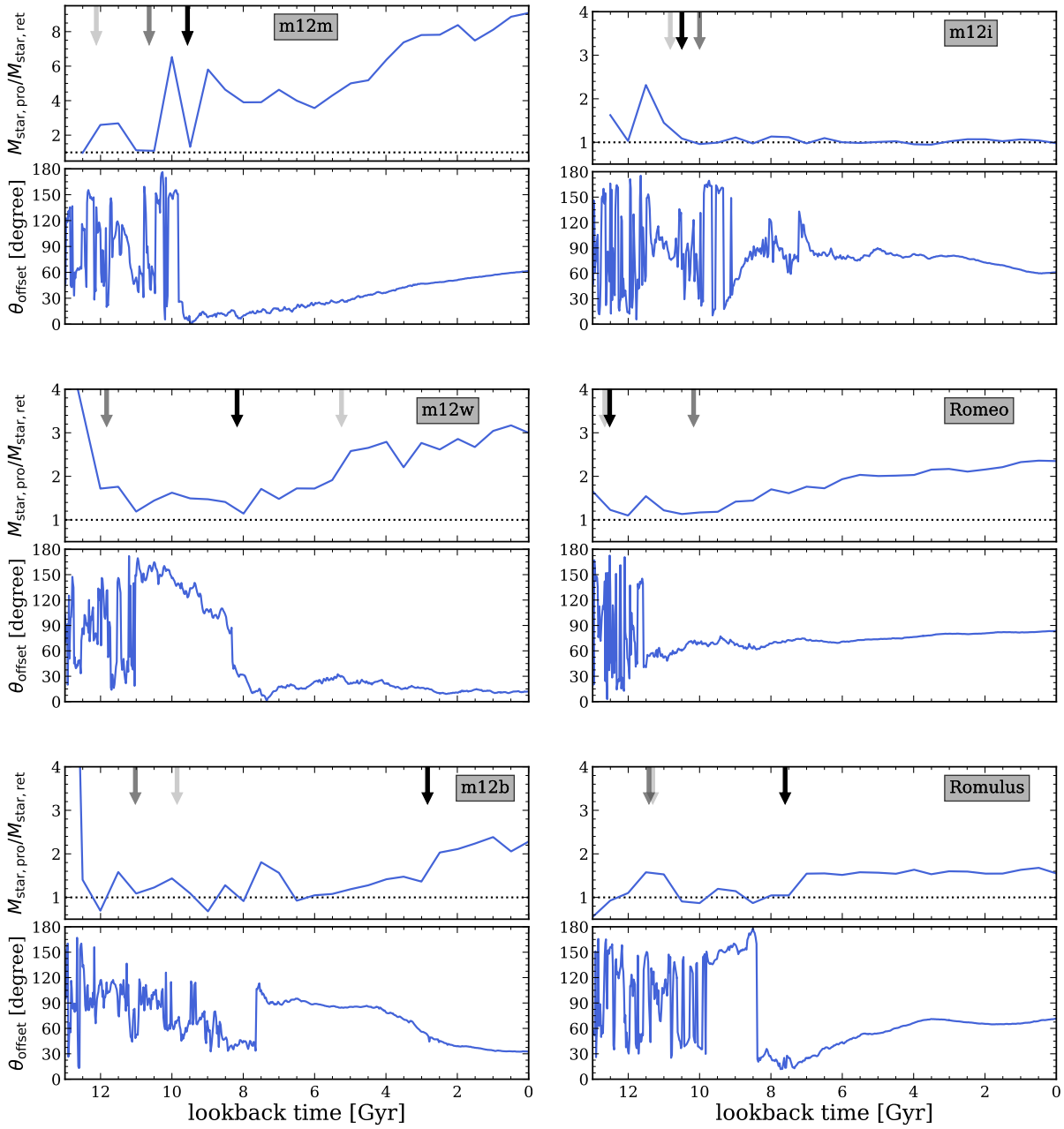

FIGURE 3.7. See the figure caption in the text.

gas-rich and drove (via torquing and gas deposition) the formation and/or prograde orientation of the host's disk, seeding the origin of prograde metal-poor stars.

We next investigate how the prograde bias evolved over time. We calculate the prograde bias at each snapshot spaced $\approx$ 500 Myr, selecting stars based on the criteria in Section 3.3.3. We

measured the orientation of the host's disk separately at each snapshot, defined according to the moment of inertia tensor of all star particles in the host. We then calculated the prograde bias at each snapshot based on the host's rotational velocity, $v_{\rm rot}$, for normalizing J_ϕ. Finally, we do not limit this analysis across time to star particles that are confined to the plane of the disk throughout their lifetimes, only those that are within the action selection windows discussed above.

Figure 3.7 (top sub-panels) shows the evolution of the prograde bias for 6 of our hosts as a function of lookback time. These hosts show a wide range of prograde biases and histories, and they are representative of the other 6 hosts that we do not show. The vertical arrows show when each primary, secondary, and tertiary merger took place (darkest to lightest, respectively).

Figure 3.7 (top left) shows m12m, the host with the strongest prograde bias, ≈ 9. As with many hosts, the prograde bias at early times was highly variable, caused by rapid merger/accretion activity and no well-defined long-lived disk. The prograde bias increased dramatically after the secondary and primary mergers, with continued gradual growth after the primary merger ~ 6 Gyr ago.

Figure 3.7 (top right) shows our one host, m12i, that has no significant prograde bias over the last 10.5 Gyr. The reason may be that all 3 mergers occurred at nearly the same time, possibly canceling out any coherent effects. Both m12m and m12i highlight the extremes of the prograde biases found in our simulated sample.

Both m12w (middle left) and Romeo (middle right) are interesting because their prograde bias steadily increased across most of cosmic time, and their the most recent merger was not their primary merger. The primary merger in m12w occurred $\sim$ 8.2 Gyr ago and significantly increased the prograde bias, while the primary merger in Romeo occurred $\sim$ 12.5 Gyr ago and had a more mild effect on the prograde bias, which subsequently decreased. Interestingly, the tertiary merger in m12w continued to drive the prograde bias upward, and more importantly, the secondary merger corresponded to the main increase for Romeo. Thus, secondary and tertiary mergers, which contributed fewer prograde metal-poor stars than the primary merger, still can effect the prograde

bias. Juliet, Louise, Remus, and m12f (not shown) are similar to Romeo, because they have gradually and monotonically increasing asymmetry biases over time.

m12b (bottom left) has the latest primary merger, ~ 3 Gyr ago, which increased the prograde bias to ≈ 2. Here, the (early) secondary and tertiary mergers did not have much of an effect. m12b shows that a galaxy can have no significant long-term prograde bias throughout most of its history, until a late merger induces one. Both Thelma and m12c (not shown) show similar behavior, with a negligible prograde bias until $\sim 6-7$ Gyr ago, which then continued to increase. In Thelma, there was no prograde bias until ~ 5 Gyr ago, however, the primary merger occurred ~ 8.5 Gyr ago and had no lasting affect on its prograde bias. The primary merger in m12c occurred 9 Gyr and did cause the prograde bias to spike to ~ 3.5, however, the secondary merger caused the prograde bias to completely vanish around 0.5 Gyr later. In both m12c and Thelma, it is likely that we do not track the mergers that cause their prograde biases to increase and become long-lived.

Finally, figure 3.7 (bottom right) shows Romulus, which had two distinct periods of significant prograde bias. The secondary and tertiary mergers induced the first period ~ 11.5 Gyr ago, but it quickly went away. The primary merger then occurred ~ 7.5 Gyr ago, which drove the prograde bias to ~ 1.5, where it remained to $z = 0$. This primary merger in Romulus contributed the highest fraction of prograde metal-poor stars compared to all other hosts (53 per cent). Both m12b and Romulus show how the timing of these mergers affects the prograde bias.

We tested several variations in how to measure the prograde bias for each host across time, and we found similar overall trends using all of them. First, we tested using the same J_ϕ selection region for each host at all snapshots, instead of scaling to each host's $v_{\rm rot}$ at each snapshot. Second, while Figure 3.7 shows the prograde bias using all metal-poor stars within the host's selection region at each snapshot, we tested following back only the star particles that are in the selection region at $z = 0$. Finally, we examined how the prograde bias evolves selecting all metal-poor stars within a radius of 15 kpc from the host, as opposed to our fiducial 'disk' spatial selection. For all of these

variations, we found the same general evolutionary trends, which reinforces that the prograde bias is a global feature of these MW/M31-mass galaxies.

3.4.4. How did the primary merger affect the orientation of the host's disk? We next investigate the relation between the primary merger's orbit and the orientation of the stellar disk of the host galaxy. In particular, we examine whether the merger was on an orbit that was aligned with the host's pre-existing disk, whether the merger torqued the direction of a pre-existing disk, and/or whether the merger helped to seed the formation and orientation of the disk.

The bottom sub-panels in Figure 3.7 show the evolution of the offset angle, $\theta_{\rm offset}$, for each host, defined as the angle between the angular momentum of the primary merger's orbit and the angular momentum of the host galaxy's disk. We store the final angular momentum vector of the primary merger just it merged into the host, and we compare it with the evolving angular momentum of the disk after the merger.

At early times, $\theta_{\rm offset}$ changed rapidly, caused by rapid accretion and mergers with many galaxies that built the host galaxy (e.g. Santistevan et al., 2020). As a result, the angular momentum of the host's (often poorly defined) disk quickly changed. All hosts in Figure 3.7 except m12b show a rapid transition from a rapidly varying $\theta_{\rm offset}$ to a settled long-lived disk orientation. This transition often coincides with one of the major mergers.

In both m12m and m12w (top left and middle left panels, respectively), just prior to the primary merger $\theta_{\rm offset}$ dipped significantly, and after the merger it went to near 0°. This means the merging galaxy was aligned with the *resultant* host's disk *after the merger*, that is, the galaxy merged into the host, deposited the metal-poor stars that remain until $z = 0$, and defined the prograde direction of the host's disk. These events occurred early and helped seed the formation of a stable long-lived stellar disk in the host. After these merger events, the disk continued to gradually change its orientation, presumably from interaction with other galaxies and/or accretion events. In m12m, the disk ends up tilted by $\approx 60°$ with respect to the primary merger orbit. However, in m12w, the disk did not rotate much, remaining at $\approx 10°$ with respect to the primary merger orbit.

Romulus (bottom right panel) shows similar behavior as m12m and m12w; starting ~ 500 Myr before the merger, the host's disk flipped from $\approx 180°$ down to$\approx 20°$ and stayed there for ~ 1 Gyr during/after the merger, after which it gradually increased to $\approx 70°$, similar to both Romeo and m12i (middle right and top right panels, respectively). Both Juliet and m12c are similar to m12m, m12w, and Romulus, because there were temporary merging events that drove $\theta_{\rm offset}$ down, while subsequent merger events caused $\theta_{\rm offset}$ to rise/rotate again.

Romeo and m12i are different from the rest of the hosts shown here, because $\theta_{\rm offset}$ remained rapidly variable both before and after the primary merger, stabilizing only $\sim 1-2$ Gyr after. The remaining hosts not shown here (m12f, Thelma, Louise, and Remus) reflect the same behavior as Romeo and m12i: rapid change in $\theta_{\rm offset}$ at early times, including before, during, and after the primary merger, but eventually the disk settleed and $\theta_{\rm offset}$ remained fairly constant at $30-140°$.

Finally, m12b (bottom left panel) is unique among our suite, because the primary merger occurred late, ~ 2.9 Gyr ago. By this time, the host's disk already formed and established itself, and because primary merger was only < 20 per cent of the stellar mass of the host, the merger only moderately torqued the orientation of the host's disk. m12b provides our single example that a *late-time* merger can drive significant prograde bias as well.

The amount of disk precession since it stabilized varies across all galaxies. The change in offset angle defined above has no distinct trend in the direction of change, that is, in some hosts $\theta_{\rm offset}$ increased over time and in others $\theta_{\rm offset}$ decreased or remained constant. The changes in $\theta_{\rm offset}$ range from $\sim 0-115°$, with a median of 25° and 68 per cent scatter of $\sim 15-35°$. We find comparable results for the change in the orientation of the disk from its $z = 0$ orientation to when it stabilized, ranging from $\sim 5-130°$, with a median of 15° and 68 per cent scatter of $\sim 10-45°$. Thus, we find moderate precession of the disk since its direction stabilized.

As we discuss in further detail in Section 3.4.5, the strength of the prograde bias does not correlate with the gas or stellar masses of the primary mergers, nor the gas or stellar mass ratios of the primary mergers to their host galaxies. At first glance, it may seem surprising that there is no

correlation of these merger/host properties with the prograde bias because these mergers largely source the prograde metal-poor population, however, as we mentioned above, they also deposit gas that contributes to the formation of the host's disk and set the prograde direction. These lacks of correlations hold true across different times (when the primary merger occurred, when the primary merger was at its peak stellar mass, 300 Myr before the primary merger) and across different ways we spatially and kinematically select gas particles in the galaxies.

We also checked whether the orbits of the primary, secondary, and tertiary mergers were aligned with each other. Specifically, we computed the orbital angular momentum vector of each merging galaxy prior its merging into the host, and we computed the offset angles between the primary and secondary, the primary and tertiary, and the secondary and tertiary. The majority of these offset angles (32 out of 36 pairs) were $\gtrsim 25°$, that is, not particularly aligned. We also checked if host galaxies with more aligned mergers have larger prograde biases, but we found no clear correlation between the alignment of the mergers and the strength of the prograde bias (p-values range from 0.21 - 0.56 for the three offset angles).

3.4.5. Correlations with galaxy properties. *Figure 3.8 caption:* The prograde bias of metal-poor ([Fe/H] < -2.5) stars in each MW/M31-mass galaxy at $z = 0$ versus various properties of its primary merger galaxy. While we do not find strong correlations for these properties, these panels indicate the range of primary mergers to highlight their diversity. **Top left**: the median offset angle between the angular momentum vector of the host's disk and the angular momentum vector of the primary merger's orbit, $\theta_{\rm offset,median}$, over 500 Myr prior to the merger. $\theta_{\rm offset,median}$ indicates how prograde the merger's orbit was with respect to the host's existing disk, with $0°$ indicating complete alignment. The prograde bias does not correlate with $\theta_{\rm offset,median}$; rather, the primary merger drove the orientation of the host's disk after the merger. **Top right**: the lookback time when the primary merger occurred, $t_{\rm lb,merger}$. Most mergers occurred $7.5 - 12.5$ Gyr ago, although m12b is an example of a late merger, only ≈ 3 Gyr ago. We find a weak negative correlation with $t_{\rm lb,merger}$, such that earlier primary mergers cause a lower prograde biases, likely because the host

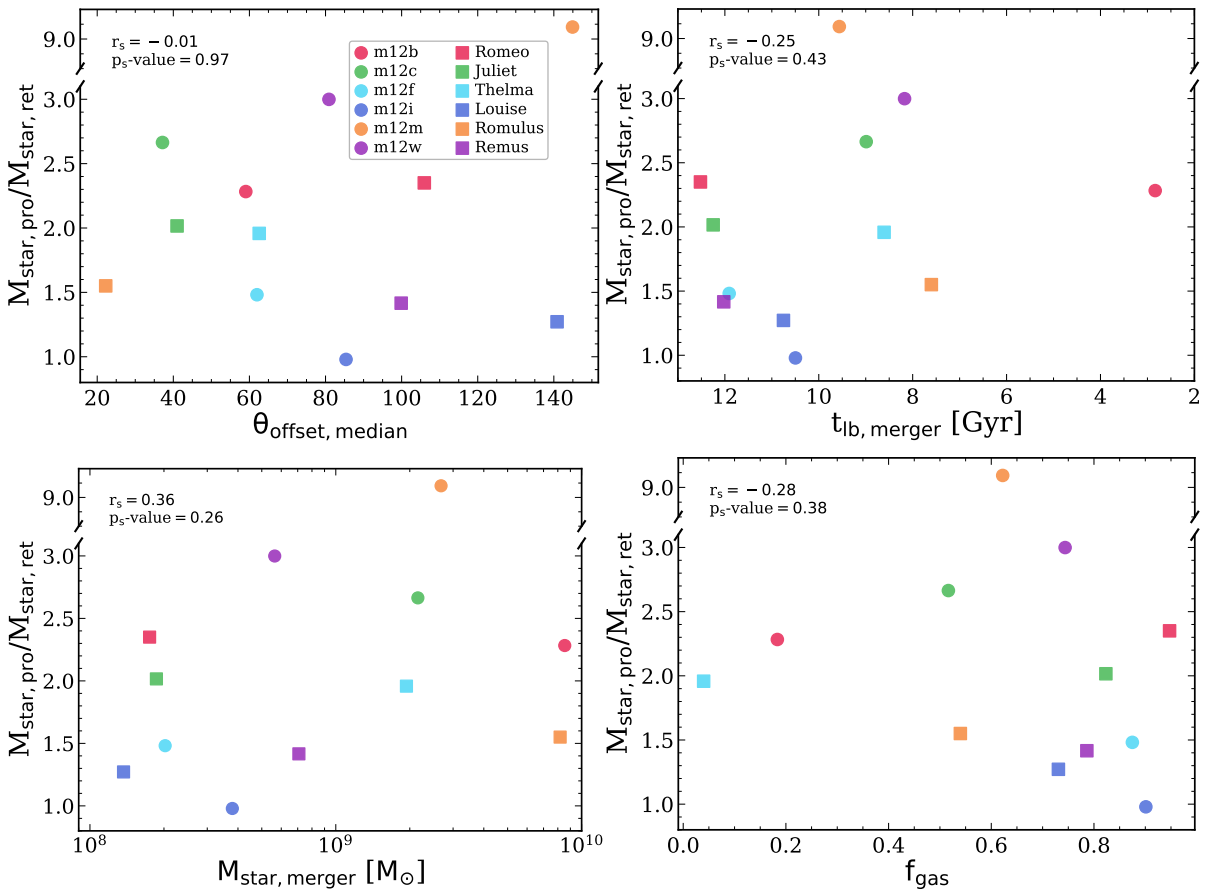

FIGURE 3.8. See the figure caption in the text.

galaxies have had more time to merge with other galaxies, which can phase-mix the stars. **Bottom left**: the stellar mass of the primary merger, $M_{\rm star,merger}$, which ranges from $10^8 - 10^{10}\,{\rm M}_\odot$, and which ranges from $0.1 - 1\times$ the stellar mass of the host at the time of the merger. The prograde bias increases weakly with $M_{\rm star,merger}$, as expected. **Bottom right**: the gas fraction of the primary merger, $f_{\rm gas,merger} = M_{\rm gas,merger}/\left(M_{\rm gas,merger} + M_{\rm star,merger}\right)$. This ranges from $\approx 0.05 - 0.95$, though most (10 of 12) primary mergers have $f_{\rm gas,merger} > 0.5$. The prograde bias decreases weakly with $f_{\rm gas,merger}$. Each panel lists the Spearman correlation coefficient and p-value: we find no strong correlation between the prograde bias and these or almost any other property that we tested (see Table 3.2). Thus, while metal-poor stars nearly ubiquitously prefer prograde orbits in our

simulations, the strength of this prograde bias has a complex dependence on formation/merger history.

Finally, we investigate correlations between properties of the MW/M31-mass host galaxies, or their galaxy mergers, and their prograde biases. Figure 3.8 shows distributions of key properties of the primary mergers and lists the Spearman correlation coefficients and p-values with respect to the prograde bias. Table 3.2 lists all of the correlations that we tested.

Given that both in-situ stars and the primary merger contribute significantly to the prograde stars, we explore whether hosts whose primary merger's orbit was more aligned with its disk orientation *before the merger* show a stronger prograde bias. Specifically, we examine the median $\theta_{\rm offset}$ over the 500 Myr before the primary merger, and Figure 3.8 (top left) shows its relation to the prograde bias at $z = 0$ for all hosts. $\theta_{\rm offset,median}$ ranges from $\sim 20 - 145°$, with Romulus having the smallest and m12m having the largest offset. In particular, m12m has both the largest pre-merger offset angle and the largest prograde bias. m12w and m12i have values in the middle of the sample, even though they have the second-most and least prograde biases, respectively. *We thus conclude that the orientation of the primary merger's orbit with the host's pre-existing disk has no significant role in driving prograde bias for metal-poor stars.*

Figure 3.8 (top right) shows the prograde bias versus the lookback time of the primary merger. Almost all primary mergers occurred > 7 Gyr ago, with the sole exception of m12b, whose merger occurred ≈ 3 Gyr ago. We find a weak correlation that hosts with an earlier primary merger tend to have a smaller prograde bias, although this correlation is weak. This correlation may arise because hosts with earlier mergers have had more time to subsequently phase-mix their star via subsequent mergers, for example.

Figure 3.8 (bottom left) shows the prograde bias versus the stellar masses of the primary merger. Because stars can start to be stripped from the merging galaxy well before coalescence, we measure the peak stellar mass of the merging galaxy throughout its history. Among our sample, the peak stellar mass of the primary merger spans $10^8 - 10^{10}\,{\rm M}_\odot$, roughly evenly in log mass. Given the

stellar masses of the host galaxies at these merger times (Santistevan et al., 2020), the primary merger galaxy was $\sim 7-95$ per cent as massive as the host during the merger. As expected, the prograde bias increases with the stellar mass of the primary merger, but the strength of this correlation is weak, as is the correlation with the the ratio of stellar masses between the primary merger and the host (see Table 3.2).

We also examine whether the gas content of the primary merger correlates with the prograde bias. The prograde metal-poor stars formed $\gtrsim 12$ Gyr ago, before the primary merger occurred, so the gas content during the merger event did not contribute to this population. Rather, the (more metal-rich) gas deposited by the merger could contribute to the formation/stabilization of the host's disk, because the merger deposited significant gas mass on a single angular momentum vector. Thus, one might expect that higher gas content in the primary merger would drive stronger prograde-ness of the deposited metal-poor stars with respect to the resultant disk. Because, gas can be stripped as the merger orbits around and coalesces with the host, we measure the gas mass of the merging galaxy 300 Myr before merging. We include all gas that is within $R_{\rm star,90}$, the radius that encloses 90 per cent of the stellar mass of the merging galaxy, and that has relative velocity $< 2 \times \sigma_{\rm star}$, the stellar velocity dispersion of the merging galaxy. Most of these primary mergers were gas-rich, bringing in between $\approx 10^8 - 10^{10}$ M$_\odot$ of gas. We then calculate their gas fractions as $f_{\rm gas,merger} = M_{\rm gas,merger}/(M_{\rm gas,merger} + M_{\rm star,merger})$.

Figure 3.8 (bottom right) shows the prograde bias versus the primary merger's $f_{\rm gas,merger}$. Most (10 of 12) primary mergers were gas-rich, with $f_{\rm gas,merger} > 0.5$, and more than half had values > 0.75. The two exceptions are m12b, which merged most recently, and Thelma, whose primary merger occurred ~ 9 Gyr ago but still was gas poor. Perhaps surprisingly, we find that more gas-rich primary mergers correspond to slightly weaker prograde bias, though again this correlation is weak. For example, m12i has the smallest prograde bias but also the second-highest $f_{\rm gas,merger}$ (≈ 0.9), while m12m has the largest prograde bias but also an intermediate $f_{\rm gas,merger}$.

TABLE 3.2. Properties that we tested for a correlation with the prograde bias of metal-poor stars at $z = 0$. Column list: property name; Spearman correlation coefficient, r_s; Spearman p-value. We calculate the mass properties/ratios at the following different times: present-day (t_0); at the time of the primary merger ($t_{\rm merger,1}$); at the time the primary galaxy merger had its peak stellar mass ($t_{\rm merger,\ peak}$); and 300 Myr before the primary merger (t_{300}).

Property	r_s	p_s value
$f_{\rm in-situ}$	-0.11	0.73
$f_{\rm merger,1}$	0.31	0.33
$M_{\rm star,\ host}(t_0)$	0.31	0.33
$M_{\rm halo,host}(t_0)$	0.10	0.76
$M_{\rm gas,host}(t_{300})$	0.23	0.47
$M_{\rm star,host}/M_{\rm halo,host}(t_0)$	0.39	0.22
$M_{\rm star,\ merger}(t_{\rm merger,\ peak})$	0.36	0.26
$M_{\rm gas,\ merger}(t_{300})$	-0.09	0.78
$M_{\rm star,\ merger}/M_{\rm star,host}(t_{\rm merger})$	0.20	0.53
$M_{\rm star,merger}/M_{\rm star,host}(t_{\rm merger,peak})$	0.06	0.85
$M_{\rm gas,\ merger}/M_{\rm gas,\ host}(t_{300})$	-0.57	0.05
$M_{\rm gas,\ total}(t_{300})$	0.18	0.57
$f_{\rm gas,\ merger}$	-0.28	0.38
$t_{\rm lb,\ merger}$	-0.25	0.43
$t_{\rm lb,\ 50}$	-0.43	0.16
$t_{\rm lb,\ 90}$	-0.28	0.38
$t_{\rm lb,\ form}$	-0.34	0.27
$N_{\rm sat}$	0.52	0.09
Primary merger ordering	0.31	0.32
$\upsilon_{\rm rot}/\sigma$	0.29	0.35
$\theta_{\rm offset,median}$	-0.01	0.97

In total, we examined correlations of the prograde bias at $z = 0$ with 21 properties of the host (both today and near the time of the primary merger) or of the primary galaxy merger, which we show in Table 3.2. Because we measure each property in Table 3.2 at different times, we calculate the Spearman correlation coefficient (r_s) and p-value at present-day (t_0), at the time of the primary merger ($t_{\rm merger,1}$), at the time the primary galaxy merger had its peak stellar mass ($t_{\rm merger,peak}$), and 300 Myr before the primary merger (t_{300}). We briefly address each below:

- $f_{\rm in-situ}$ and $f_{\rm merger,1}$: While $f_{\rm merger,1}$ does positively correlate with the prograde bias, the strength of its correlation is weak, indicating additional dependence, such as time and gas-richness of the merger. Importantly, $f_{\rm in-situ}$ shows no meaningful correlation with the prograde bias.
- The ratio of stellar mass of the primary merger and the host does not correlate significantly with prograde bias. We tested measuring this ratio both at the time when the primary merger reached its peak stellar mass ($M_{\rm star,merger}/M_{\rm star,host}(t_{\rm merger,peak})$, and at the time immediately before merging into the host galaxy ($M_{\rm star,merger}/M_{\rm star,host}(t_{\rm merger})$.
- $M_{\rm gas,merger}/M_{\rm gas,host}(t_{300})$: the ratio of the gas mass in the primary merger to that in the host galaxy. We explored four ways of measuring gas masses, including 300 Myr or 600 Myr before the primary merger, using only cold gas ($T < 2 \times 10^4$K), or measuring gas out to $2\times R_{90}$. Interestingly, this ratio 300 Myr before the merger is the most significant correlation that we find (p-value = 0.05). However, we caution the reader to over-interpreted this result, because the strength of this correlation varies greatly with how we measure gas masses (p-values = 0.23-0.78). In particular, the gas mass of the primary merger is relatively robust across our methods, but the gas mass of the host galaxy varies more significantly, which may be because of ambiguity in mass association during or just before a merger.
- Host mass: $M_{\rm star,host}(t_0)$, $M_{\rm halo,host}(t_0)$, $M_{\rm gas,host}(t_{300})$, and $M_{\rm star,host}/M_{\rm halo,host}(t_0)$, correlate at most weakly with prograde bias. The strongest correlation is the current ratio of stellar mass to dark-matter halo mass.
- $M_{\rm gas,total}(t_{300})$: the total gas mass, summing that of the primary merger and the host galaxy, does not correlate significantly with prograde bias.
- $t_{\rm lb,50}$, $t_{\rm lb,90}$, and $t_{\rm lb,form}$: we investigated three metrics of host galaxy formation: the time when the galaxy formed 50 per cent of its final mass, the time when it reached 90 per cent of its final mass, and the time when it transitioned from having primarily ex-situ to primarily in-situ stellar mass. All of these correlate negatively with prograde bias, however, the

significance of these correlations remains weak. $t_{\rm lb,50}$ correlates most strongly, perhaps because it correlates best with typical times of the most important mergers. Despite differences in the formation histories between galaxies in LG-like pairs, and galaxies in isolated environments mentioned in Santistevan et al. (2020), and the small positive correlation between the primary merger times ($t_{\rm lb,merger}$) and host galaxy formation times ($t_{\rm lb,form}$; p-value=0.05), we found no significant differences in the metal-poor, prograde biases between LG-like and isolated host galaxies, which suggests prograde bias does not correlate with host galaxy formation time.

- $N_{\rm sat}$: the number of satellite dwarf galaxies ($M_{\rm star} > 10^5$ M$_\odot$) within the host halo at $z = 0$ is curiously one of the strongest correlation that we find (p-value = 0.09). It remains unclear whether this is physically meaningful or a statistical fluke among our modest sample size. Taken at face value, it would imply that M31 has an even stronger prograde bias than the MW.
- Primary merger ordering: we compare the relative timing of the primary merger to the secondary or tertiary mergers, that is, whether the primary merger occurred first, second, or third, which we define as the primary merger ordering. We find only a weak correlation, such that hosts whose primary merger happened the most recently out of the 3 have a somewhat stronger prograde bias.
- $\upsilon_{\rm rot}/\sigma$: the ratio of the rotational velocity to the velocity dispersion of stars in the host at $z = 0$, $\upsilon_{\rm rot}/\sigma$ is a kinematic metric of 'diskiness'. We calculated both $\upsilon_{\rm rot}$ and σ using all stars in the disk, however, we also calculated the ratio using only stars that are metal-poor (threshold from $-3.5 <$ [Fe/H] < -0.5). We find a marginal increase in prograde bias with $\upsilon_{\rm rot}/\sigma$, but it is weak.

3.5. Summary & Discussion

3.5.1. Summary. We investigated the prevalence and origin of the preference of metal-poor stars to be on prograde disk orbits using 12 MW/M31-mass galaxies from the FIRE-2 simulation

suite. We reiterate the questions that we articulated in the introduction, and we summarize our answers to them:

i) *How commonly do MW/M31-mass galaxies show a preference for prograde motions in their metal-poor stars, as observed in the MW, and what is the range in strength of this prograde bias?*
   - 11 of 12 of our MW/M31-mass galaxies show a preference for prograde orbits among metal-poor stars (Figure 3.1).
   - Our simulations predict that this prograde bias is a general feature of MW/M31-mass galaxies throughout most of their history, and its presence does not depend significantly on the way that we spatially or kinematically select metal-poor stars (Figures 3.2, 3.3, 3.5).
   - The present-day prograde biases for our sample range from $0.98-9.14$, with a median value of 2.0 (Table 3.1).
   - We find no significant differences in prograde bias between LG-like and isolated host galaxies (Figure 3.2).

ii) *How does this prograde bias depend on the metallicity, distance, and/or age of the stars?*
   - We find little-to-no dependence of the prograde bias on the iron abundance of stars at [Fe/H] $\lesssim -1$ (Figure 3.2).
   - Our results broadly agree with recent observations of the MW from the Pristine and H3 surveys (Figures 3.2 & 3.3).
   - We find typically a similar prograde bias, $\sim 2-2.5$, for stars at distances $\gtrsim 20$ kpc in the stellar halo (Figure 3.4).
   - 11 of 12 of our galaxies have a prograde bias for arbitrarily old stars ($\gtrsim 13.5$ Gyr) regardless of metallicity, and all hosts have a significant prograde bias for stars $\lesssim 12.5$ Gyr old (Figure 3.5).

iii) *What process(es) cause this prograde bias among metal-poor stars?*

- We find no large difference in the prograde bias for stars that formed in-situ versus ex-situ (Figure 3.2). The sample of prograde metal-poor stars (both in-situ and ex-situ) almost entirely formed $\gtrsim 12$ Gyr ago, prior to the primary mergers.
- A *single* galaxy merger is typically the most important cause of the prograde bias, with additional but sub-dominant contributions from in-situ star formation and lesser mergers (Figure 3.6). This primary merger typically contributed ~ 24 per cent of all prograde metal-poor stars, while the fraction that formed in-situ is ~ 16 per cent. Combining in-situ stars with those from the top 3 mergers accounts for ~ 70 per cent, on average (Table 3.1).
- We do not find any significant correlation in orientation between the primary merger's orbit and any pre-existing disk in the host (Figure 3.7).
- The primary merger occurred typically $7 - 12.5$ Gyr ago and typically was gas rich ($f_{\rm gas,merger} \gtrsim 0.5$). It thus deposited a significant population of old metal-poor stars and significant gas into the host on the same orbital vector, which typically seeded/shaped the formation of a long-lived disk in the host, giving rise to metal-poor stars preferentially on a prograde disk orbit (Figures 3.7 & 3.8).
- In 3 of our simulated hosts, the host stellar disk orientation became settled/long-lived within ~ 500 Myr of 1 of the 3 mergers we tracked, and in two cases, the orientation became settled, but it subsequently changed via other mergers (Figure 3.7). This suggests that mergers can be a dominant factor in the formation of stellar disks.

iv) *Do any properties of the MW/M31-mass galaxy or their galaxy mergers correlate with this prograde bias?*

- We explored correlations with 21 properties of the host galaxy or primary merger, but we find few clear correlations with the strength of the prograde bias (Table 3.2). This indicates that the strength of the prograde bias has a complicated dependence on galaxy formation/merger history, which limits our immediate interpretation of how to

translate recent observations of the prograde bias for the MW into a robust constraint on its formation history.

3.5.2. Discussion. A few galaxy mergers predominantly drive the growth of the metal-poor component of MW/M31-mass galaxies/halos. For example, using a suite of dark-matter-only simulations of MW-mass halos and applying abundance matching to assign stellar mass to (sub)halos, Deason et al. (2016) found that most of the accreted stellar mass in the stellar halo comes from only 1-2 dwarf galaxies with $M_{\rm star} = 10^8 - 10^{10}\,{\rm M}_\odot$. Similarly, in Santistevan et al. (2020) we used the same 12 simulated galaxies in this paper and found that stars that form ex-situ (in galaxies other than the most massive progenitor) and accrete into the host galaxy largely come from the most massive progenitor galaxies, and that this is true for progenitors at all redshifts. Because just a few galaxies are responsible for most of the accreted (metal-poor) population, they leave a kinematic imprint, as we have shown. Accreted substructures populate our own galaxy, such as Gaia-Enceladus (Helmi et al., 2018; Belokurov et al., 2018), the Sagittarius stream (Ibata et al., 1994; Newberg et al., 2003; Majewski et al., 2003), Sequoia (Barbá et al., 2019; Myeong et al., 2019; Matsuno et al., 2019), and a few others (e.g. see Naidu et al., 2020).

Although stars in Gaia-Enceladus are on primarily radial orbits, they may have some net rotation. Helmi et al. (2018), Belokurov et al. (2018), Naidu et al. (2020), and Naidu et al. (2021) all suggest that even though the mean rotational velocity of Gaia-Enceladus' stars is close to 0, it may exhibit a small degree of retrograde motion. However, this is further complicated by the possible correlation of Gaia-Enceladus with other substructures that populate the MW's stellar halo. Naidu et al. (2020) lists the currently known substructures, and details 3 prograde structures (Sagittarius & Helmi streams, Wukong) and 4 retrograde structures (Arjuna, Sequoia, I'itoi, Thamnos). It remains unclear whether or not Thamnos, Sequoia, or I'itoi are associated with Gaia-Enceladus, or with each other. Ignoring Gaia-Enceladus stars in their sample, Naidu et al. (2020) report that the number of prograde to retrograde stars is $\sim 3/2$ within the stellar halo, which qualitatively agrees with our results, though is weaker than our median value in Figure 3.4. However, the Sagittarius

stream dominates this signal in the MW, and when Naidu et al. (2020) do not include Sagittarius stars, they find no prograde bias in the stellar halo.

Because the MW shows a clear prograde bias for metal-poor stars, at least near its disk region, then the results above, combined with our results, imply one or more of the following: (i) one or more of the mergers above imparted a prograde bias to the MW, but their stars that populate the stellar halo (beyond the disk region) have since radialized; (ii) the prograde bias in the MW originated from a yet-to-be discovered merger, perhaps whose remnant stars populate the inner halo; (iii) the prograde bias in the MW is entirely in-situ and not associated with any merger. However, our results suggest that the latter scenario (alone) is rare.

In their analysis of the prograde bias in the H3 survey, Carter et al. (2021) suggest that metal-poor stars likely came from either (1) a combination of in-situ and ex-situ star formation during the early assembly and formation of the MW, which is broadly similar to what we find, or (2) late accretion of dwarf galaxies on prograde orbits with respect to the disk. However, because the stars in the MW have not been kinematically heated enough to change their prograde orbits, the authors claim that the angular momentum of the MW has been in place for at least 12 Gyr. From the bottom sub-panels of Figure 3.7, we see find changes in $\theta_{\rm offset}$ at early times, with a sharp transition, after which the angle slowly precesses over time. This transition marks when the stellar disk orientation stabilized in the host, and although the disk can precess afterwards, $\theta_{\rm offset}$ remains relatively stable. We find a range of times when this transition occurs, across $\approx 5 - 11.5$ Gyr ago. Romeo is the host galaxy that has the oldest transition akin to the MW, at ~ 11.5 Gyr ago. However, in most of our hosts, this transition occurs more recently, which suggests that the MW may be uncommon in this regard if the angular momentum of the disk has been stable for the last 12 Gyr. If instead we compare the $\hat{z}$ orientation of the disk at $z = 0$ to the orientation at early time, we see the same qualitative behavior in Figure 3.7: rapid changes likely driven by early assembly, then a sharp transition to a more gradual change from disk precession. Changes in both $\theta_{\rm offset}$ and in the disk's

orientation, from the time that these rapid changes settle to present-day, vary between $\sim 0 - 130^\circ$ and have similar median values of 25° and 15°, respectively.

Naidu et al. (2021) recently aimed to reconstruct the merger of Gaia-Enceladus with the MW. By comparing observational data from the H3 survey with N-body simulations, the authors constrained the orbital properties of the merger, such as the inclination, circularity, and direction of the orbit, as well as the stellar mass and size of the galaxy. Interestingly, the authors conclude that Gaia-Enceladus was on an initial retrograde orbit and subsequently was radialized over time via dynamical friction, but they also show the mean angular momentum of Gaia-Enceladus stars as a function of galacto-centric distance. Figure 15 of Naidu et al. (2021) suggests that Gaia-Enceladus stars within $d_{\rm host} < 25$ kpc are only slightly retrograde today, but beyond 25 kpc the debris becomes increasingly more retrograde. One caveat of their analysis is that the authors do not account for the amount of disk precession over time; we see precession between $5 - 130^\circ$ from the time the disk stabilizes to its $z = 0$ orientation. The results from Carter et al. (2021) suggested that the stellar halo stars in the H3 survey show a prograde bias, similar to our results. Our Figure 3.4 also shows that the stellar halo stars in our simulated galaxies have a *prograde* bias that increases to $\sim 2 - 2.5$ at $d_{\rm host} = 45$ kpc. However, an important clarification to make is that we show *all* stars in Figure 3.4 (and in Figures 3.2-3.5), not just stars in a single merger. This motivates future work to more carefully test our results (and those in Carter et al. (2021)) in the context of individual merger remnants, as in Naidu et al. (2021).

Although mergers are violent processes, they can result in the formation of a disk (e.g. Springel & Hernquist, 2005; Robertson et al., 2006; Sparre & Springel, 2017; Tapia et al., 2017) or change the properties of existing disks (e.g. Toomre & Toomre, 1972; Bournaud et al., 2004, 2005; Hopkins, 2009). For example, following a series of papers aimed at understanding disk formation through major mergers (Athanassoula et al., 2016; Rodionov et al., 2017; Peschken et al., 2017), Peschken et al. (2020) analyzed galaxies in the Illustris simulation (Vogelsberger et al., 2014) and showed that even though 38 of them experienced a $\gtrsim$ 4:1 merger since $z \sim 1.5$, they eventually formed

disks by $z = 0$. The most important factor in determining the final outcome of such a merger is the amount of gas, given that stars can form from this gas after the merger to build a disk, while the older pre-existing stars can be heated during the merger to form a thick-disk/halo/ellipsoidal component in the galaxy. In our simulated sample, 300 Myr prior to the primary mergers, the gas mass ratios of the primary merging galaxy to the host span values from ~ 0.002 in Thelma, to as high as ~ 1.6 in Louise. Similarly, the total gas masses (primary merging galaxy + host) 300 Myr prior to the primary mergers span values from $\sim 5 \times 10^8 - 4 \times 10^{10}\,\mathrm{M}_\odot$. In 3 of our hosts (m12m, m12w, and Romulus), the orientation of the disk settles from the rapid variations caused by early accretion/mergers within $\approx 500\,\mathrm{Myr}$ of the three mergers we investigated here, suggesting that these mergers drove the formation and prograde orientation of the disk. In 2 other hosts (m12c and Juliet), 1 of the 3 mergers caused the orientation of the disk to settle, but other subsequent mergers caused the disk to then change orientation again. Our results qualitatively agree with Peschken et al. (2020) and suggest that mergers were a dominant mode of promoting the transition to thin-disk morphology, with much of the gas mass mentioned above likely contributing to subsequent star formation in the disk region.

Recently, Sestito et al. (2021) performed a similar study as ours using 5 of the NIHAO-UHD cosmological zoom-in simulations of MW-mass galaxies (Buck et al., 2020). Building on their previous work, they found that the simulated galaxies also show a prevalence for a prograde bias. The authors also showed that a rotating spheroid of metal-poor stars, akin to a halo-like or pressure supported distribution of stars from the early accretion phase, cannot explain the origin of the prograde bias in a galaxy. Rather, the early accretion and mergers of progenitor galaxies is the dominant source of metal-poor planar stars, with a smaller additional contribution from later accretion of satellite galaxies on prograde orbits that get dragged and disrupted into the proto-disk. Furthermore, they suggest that $\sim 88 - 93$ per cent of metal-poor stars ([Fe/H] < -2) in those simulations are older than $\sim 12\,\mathrm{Gyr}$, where the retrograde planar population traces the early

accretion of the main galaxy and the prograde planar population better traces the full history. Our results qualitatively agree with this picture, and we discuss the similarities and differences below.

Both our analysis and Sestito et al. (2021) find that this population of metal-poor stars within the disks of MW-mass galaxies has a preference of prograde motion as opposed to retrograde. Sestito et al. (2021) find a prograde bias in all 5 NIHAO-UHD simulated galaxies, and we find a prograde bias in 11 of our 12 galaxies. This reinforces that the prograde bias for metal-poor stars is a nearly (but not completely) ubiquitous feature of MW/M31-mass galaxies. Our sample of 12 host galaxies is larger than that in Sestito et al. (2021), which is likely why we find one host without a prograde bias. The strengths of the prograde biases differ significantly between our analysis and theirs. Sestito et al. (2021) find prograde biases of $\sim 3.5 - 9.1$, higher than the MW, while our simulations have $0.98 - 9.1$, with a median value nearly identical to the MW. In particular, 11 of our hosts have prograde biases $\lesssim 3$, below the smallest prograde bias in Sestito et al. (2021), which may indicate the importance on these results from differences in physical and/or numerical implementations between our FIRE-2 and the NIHAO-UHD simulations.

We qualitatively agree about the origins of prograde metal-poor stars. These star particles primarily came from a number of mergers, both from the early accretion and late growth phases. Even though a non-negligible fraction of these stars form in-situ ($5-30$ per cent) in our simulations, they are too old to have formed in a disk-like component within the host. In Figure 5 of Sestito et al. (2021), massive satellite galaxies ($M_{\rm star} \gtrsim 10^8\,{\rm M}_\odot$) merged into the host on orbits aligned with the angular momentum of the stellar disk around the time the host galaxy formed 25 per cent of its final mass (t_{25}). At these times, they report that $\sim 37 - 90$ per cent of the prograde planar stars are already in the host galaxy from the early assembly phase. However, from the bottom sub-panels in Figure 3.7, we show that the orientation of the proto-disk is not well defined until the primary merger (or another lesser merger) occurs, implying that the primary merger helped to seed the formation and set the orientation of the host's disk by depositing a significant supply of gas on a coherent angular momentum vector. Thus, this means that the primary merger determined the

prograde direction in the host, and that these progenitor galaxies do not merge into the host on orbits aligned with the disk because the disk was not well-defined before this. Our answer to the origin of prograde metal-poor stars in the disks of MW-mass galaxies is then most similar to explanation (iii) from Sestito et al. (2019): these stars formed inside one or more progenitor galaxies that merged into the MW-mass host as it was forming.

As a final comparison, we note important differences between our analysis and Sestito et al. (2021). The authors do not explore the stability of the stellar disk orientations in their Figure 5 at t_{25}. Therefore, it is unclear what effect, if any, the satellite progenitor galaxies had on the orientation of the host. Although they quantify the growth of the accreted stellar mass of the host galaxy, they do not examine the gas mass of the merging galaxies, which plays a large role in shaping the disk. In this paper, we only focus on the evolution of the ratio of the stellar mass of prograde to retrograde stars, but Sestito et al. (2021) suggests that 74 – 90 per cent of the retrograde planar population are primarily accreted before t_{25}, while the prograde population is accreted throughout most of cosmic time. This paper also investigated several properties of the host galaxy to look for possible correlations with the strength of its prograde bias. Finally, we do not analyze other regions in kinematic space, such as the high eccentricity feature in Sestito et al. (2021).

Evidence for the prograde, metal-poor population in the MW disk is further supported by a recent study involving stars located in the stellar halo and thick disk from the *SkyMapper Survey for extremely metal-poor stars* (Da Costa et al., 2019). In a sample of 475 stars with iron abundances in the range of $-6.5 \lesssim$ [Fe/H] $\lesssim -2$, Cordoni et al. (2020) reported that $\sim$ 21 per cent of their sample have disk-like orbits confined to 3 kpc of the disk's midplane. The authors suggest that this sub-sample of metal-poor stars with disk-like orbits consists of a high-eccentricity and low-eccentricity population, where the low-eccentricity population are on primarily prograde orbits (defined as $J_\phi > 0$) with a prograde bias of ~ 3.1. Cordoni et al. (2020) postulate that the prograde and retrograde low-eccentricity stars confined to $|Z| < 3$ kpc likely have different origins: the prograde stars possibly formed in the thick disk of the MW, while the retrograde were accreted

from a disrupted satellite galaxy. These results broadly agree with what we presented here, however, it is likely that some of the prograde, low-eccentricity stars formed both in-situ *and* were accreted early in the MW's formation history.

Similar studies focused on old, metal-poor stars in the MW also halo have investigated their origin and kinematics. By selecting stars within 3 kpc of the Sun with halo-like kinematics, Bonaca et al. (2017) found that there are an excess of metal-rich halo stars ([Fe/H] > -1) with prograde orbits compared to the metal-poor halo stars ([Fe/H] < -1) which have no orbital preference. The authors compared this with one of the simulations in our sample, m12i, and found similar results. They found that the two halo populations also have different origins, with the metal-rich stars primarily forming in-situ, and the metal-poor population accreting from mergers (ex-situ). A similar study using the APOSTLE simulations by Starkenburg et al. (2017a) suggests that the fractions of metal-poor ([Fe/H] < -2.5) and old ($t_{\rm form} < 0.8\,$Gyr) stars increase for larger galactocentric radii, with ~ 60 per cent of the stars orbiting outside of the solar circle. We did not examine the prograde bias of the halo region in our simulations, but rather, we compared how the prograde bias evolution changed when selecting *all* metal-poor ([Fe/H] < -2.5) stars within 15 kpc, radially. We saw the same qualitative evolution of the prograde biases as in the top sub-panels of Figure 3.7. This is not a strict 'halo selection' of stars, but given that we showed that the prograde bias is a general feature, independent of the kinematic and spatial selection windows we examined, we do not expect our results to change significantly by making a more strict selection of halo stars. Similarly, El-Badry et al. (2018b) investigated the spatial distribution and dynamics of the oldest stars ($z_{\rm form} \gtrsim 5$) in MW-mass galaxies and found that the majority formed ex-situ. The authors found that metal-rich stars ([Fe/H] > -1) typically have disk-like orbits, while the more metal-poor stars reflect an isotropic distribution. Furthermore, when only selecting the oldest stars ($z_{\rm form} \gtrsim 5$), the in-situ and ex-situ populations in m12f have nearly identical velocity distributions, however, both populations do show a slight prograde bias, which are centered at $v_\phi > 0\,{\rm km\,s^{-1}}$.

Instead of selecting stars based on their distributions in velocity space, as in a Toomre diagram, one recent study tested a Gaussian mixture model to deconstruct the different components of the MW using the star's velocities and iron abundances (Nikakhtar et al., 2021). The authors used data from APOGEE (Majewski et al., 2017) and Gaia DR2 (Gaia Collaboration et al., 2018), and created mock catalogs using 3 of the simulations in our study (m12f, m12i, m12m), and showed that the MW is best described by five different components: the stellar halo, the thin disk, the metal-rich thick disk, and a 2-component metal-poor thick disk. Nikakhtar et al. (2021) suggest that the 3-component thick disk originates from a combination of kinematic heating (via gravitational interactions) and radial migration of early forming stars: we show that a significant fraction of the most metal-poor stars in the disk come from multiple mergers during the galaxy's early assembly. In a different approach, using the same 12 simulated galaxies as in our sample, Yu et al. (2021) differentiated thin disk and thick disk stars based on their circularity (angular momenta relative to the angular momentum of a circular orbit) and suggest that the transition from thick disk to thin disk formation correlates with the transition from early, bursty star formation to steady, near constant star formation. This transition occurred some time between $\sim 0.6 - 6.5$ Gyr ago, and galaxies that had earlier transitions often have older thick disk stars and more prominent thin disks.

One interesting avenue for future work is to investigate the correlation of the prograde bias with larger-scale structure. To first order, we found no distinct differences in the metal-poor prograde bias between isolated and LG-like environments, which suggests that the prograde bias is not inherent in LG-like environments only, as in the MW.

Acknowledgements

We greatly appreciate interesting and fruitful discussions with both Federico Sestito and Nicolas Martin, as well as Courtney Carter and Charlie Conroy. We also express our gratitude for their sharing of observational data. Finally, we wish to thank the reviewer, Ted Mackereth, for offering many great suggestions to strengthen the paper.

IBS, AW, and JS received support from NASA through ATP grants 80NSSC18K1097 and 80NSSC20K0513; HST grants GO-14734, AR-15057, AR-15809, and GO-15902 from STScI; a Scialog Award from the Heising-Simons Foundation; and a Hellman Fellowship. AW performed this work in part at KITP, supported by NSF grant PHY-1748958. CAFG was supported by NSF through grants AST-1715216 and CAREER award AST-1652522; by NASA through grant 17-ATP17-0067; by STScI through grant HST-AR-16124.001-A; and by a Cottrell Scholar Award and a Scialog Award from the Research Corporation for Science Advancement. RES acknowledges support from NASA grant 19-ATP19-0068, NSF grant AST-2009828, and HST-AR-15809 from the Space Telescope Science Institute (STScI), which is operated by AURA, Inc., under NASA contract NAS5-26555. We ran simulations using: XSEDE, supported by NSF grant ACI-1548562; Blue Waters, supported by the NSF; Pleiades, via the NASA HEC program through the NAS Division at Ames Research Center.

This paper used various python packages including NumPy (van der Walt et al., 2011), SciPy (Virtanen et al., 2020), and Matplotlib (Hunter, 2007), as well as NASA's Astrophysics Data System.

CHAPTER 4

Orbital dynamics and histories of satellite galaxies around Milky Way-mass galaxies in the FIRE simulations

Published as Isaiah B. Santistevan, Andrew Wetzel, Erik Tollerud, Robyn E. Sanderson, Jenna Samuel in Monthly Notices of the Royal Astronomical Society, Volume 518, Issue 1, January 2023, Pages 1427-1447, https://doi.org/10.1093/mnras/stac3100

4.1. Abstract

The orbits of satellite galaxies encode rich information about their histories. We investigate the orbital dynamics and histories of satellite galaxies around Milky Way (MW)-mass host galaxies using the FIRE-2 cosmological simulations, which, as previous works have shown, produce satellite mass functions and spatial distributions that broadly agree with observations. We first examine trends in orbital dynamics at $z = 0$, including total velocity, specific angular momentum, and specific total energy: the time of infall into the MW-mass halo primarily determines these orbital properties. We then examine orbital histories, focusing on the lookback time of first infall into a host halo and pericentre distances, times, and counts. Roughly 37 per cent of galaxies with $M_{\rm star} \lesssim 10^7\,{\rm M}_\odot$ were 'pre-processed' as a satellite in a lower-mass group, typically $\approx 2.7\,{\rm Gyr}$ before falling into the MW-mass halo. Half of all satellites at $z = 0$ experienced multiple pericentres about their MW-mass host. Remarkably, for most (67 per cent) of these satellites, their most recent pericentre was not their minimum pericentre: the minimum typically was ~ 40 per cent smaller and occurred $\sim 6\,{\rm Gyr}$ earlier. These satellites with growing pericentres appear to have multiple origins: for about half, their specific angular momentum gradually increased over time, while for the other half, most rapidly increased near their first apocentre, suggesting that a combination of

a time-dependent MW-mass halo potential and dynamical perturbations in the outer halo caused these satellites' pericentres to grow. Our results highlight the limitations of idealized, static orbit modeling, especially for pericentre histories.

4.2. Introduction

The satellite galaxies around the Milky Way (MW) and M31 in the Local Group (LG) are unique systems in that we can measure both resolved stellar populations and full orbits to derive orbital histories. Key answerable questions regarding satellite formation histories include: What are their orbits today and what were their orbital histories? When did they first become satellites? How many were satellites of a lower-mass group, like the Large Magellanic Cloud (LMC; Kallivayalil et al., 2018; Patel et al., 2020; Pardy et al., 2020), and when did they become satellites of such groups? How close have they orbited to the MW and M31? Understanding the answers to these questions within the LG will also improve our understanding of satellite evolution in systems beyond the LG.

Thanks to numerous studies (for example Kallivayalil et al., 2013, 2018; Fritz et al., 2018b; Gaia Collaboration et al., 2018; Patel et al., 2020) and HST treasury programs (for example, GO-14734, PI Kallivayalil; GO-15902, PI Weisz), we now have or soon will have 3D velocity information for the majority of the satellites in the LG, with continued higher precision (for example, see updated values between McConnachie, 2012; Fritz et al., 2018a, and references therein). With the advent of new data coming from studies such as the Satellites Around Galactic Analogs (SAGA) survey (Geha et al., 2017; Mao et al., 2021), which focus on measuring properties of satellites of MW-mass galaxies beyond the LG, understanding general orbital/infall histories of satellites is imperative. Given this rich data, combined with observationally informed estimates of the MW's gravitational potential (for example Bland-Hawthorn & Gerhard, 2016; McMillan, 2017; Li et al., 2020a; Deason et al., 2021; Correa Magnus & Vasiliev, 2022), one can constrain the orbital histories of satellites. However, a key challenge is understanding limitations and biases from assuming a static, non-evolving potential (for example Klypin et al., 1999; D'Souza & Bell, 2022).

One of the most important events in a satellite galaxy's history is when it first orbited within the MW-mass halo, or any more massive halo. After falling into a more massive host halo, a satellite galaxy temporarily can orbit outside of the host's virial radius (a 'splashback' satellite); typically such satellites remain gravitationally bound and orbit back within the host's virial radius (for example Ludlow et al., 2009; Wetzel et al., 2014). As a satellite orbits within the host halo, its hot gas can quench the satellite's star formation through ram-pressure stripping of gas (for example Gunn & Gott, 1972; van den Bosch et al., 2008; Fillingham et al., 2019; Rodriguez Wimberly et al., 2019; Samuel et al., 2022b). Not only can a lower-mass galaxy fall into a MW-mass halo, but it can become a satellite of another intermediate-mass galaxy before falling into the MW-mass halo, called 'pre-processing' (for example D'Onghia & Lake, 2008; Li & Helmi, 2008; Wetzel et al., 2015; Deason et al., 2015; Kallivayalil et al., 2018; Jahn et al., 2019; Patel et al., 2020). This process can suppress and even quench star formation in the low-mass galaxy before it falls into its MW-mass halo (for example Jahn et al., 2022; Samuel et al., 2022b). If satellites fell into a MW-mass halo via a group, they should have similar orbits, at least for one or a few orbital timescales, with broadly similar orbital angular momenta and energy (for example Sales et al., 2011; Deason et al., 2015; Sales et al., 2017; Pardy et al., 2020). Some theoretical studies suggest that no current satellites of the MW were satellites during the epoch of reionization (for example Wetzel et al., 2015; Rodriguez Wimberly et al., 2019) at $z \gtrsim 6$, thus, if the satellites quenched at $z \gtrsim 6$, the host environment could not have quenched them, such that the effects of the host environment and cosmic reionization are separable, in principle.

Many works have studied infall histories and orbital properties of simulated satellite galaxies of MW-mass halos (for example Slater & Bell, 2013; Wetzel et al., 2015; Li et al., 2020b; Bakels et al., 2021; D'Souza & Bell, 2021; Robles & Bullock, 2021; Ogiya et al., 2021), sometimes with the intent of deriving properties of observed satellites of the MW (for example Rocha et al., 2012; Fillingham et al., 2019; Miyoshi & Chiba, 2020; Rodriguez Wimberly et al., 2022). However, many such previous works used dark matter only (DMO) simulations, and the inclusion of baryonic

physics is critical to model accurately the satellite population (see Brooks & Zolotov, 2014; Bullock & Boylan-Kolchin, 2017; Sales et al., 2022).

One important process that affects the orbital evolution of satellites is dynamical friction. As satellites orbit within a host halo, they induce an over-density of dark matter behind them, which causes a drag force called dynamical friction that slows their orbits and can lead them to merge with the host galaxy (Chandrasekhar, 1943; Ostriker & Tremaine, 1975). Dynamical friction is more efficient when the satellite and host galaxy or halo are of similar mass (for example Boylan-Kolchin et al., 2008; Jiang et al., 2008; Wetzel & White, 2010). Upon accretion, massive satellites can also induce global perturbations within the larger MW-mass galaxy, which also can affect the orbits of less-massive satellites (for example Weinberg, 1986, 1989; Colpi et al., 1999; Tamfal et al., 2021). Furthermore, because the dark matter in the satellite galaxy is tidally stripped as it orbits throughout the host halo, this stripped material further can slow the satellite (Miller et al., 2020). One way to parameterize the efficiency of dynamical friction is via the time from first infall it takes a satellite to merge into its host galaxy/halo. For example, Wetzel & White (2010) approximated this merging time as

$$t_{\rm merge} \approx C_{\rm dyn} \frac{M_{\rm host}/M_{\rm sat}}{\ln(1 + M_{\rm host}/M_{\rm sat})} t_{\rm Hubble} \tag{4.1}$$

where $M_{\rm host}$ is the total mass of the host halo, $M_{\rm sat}$ is the total mass of the smaller satellite halo, $t_{\rm Hubble} = H^{-1}(z)$. They found $C_{\rm dyn} \approx 0.2 - 0.3$ agrees well with the results of a large-volume cosmological DMO simulation. This implies that for a satellite to merge within a Hubble time (the age of the Universe) the halo mass ratio between the host and a satellite must be closer than about $1:20$.

Because dynamical friction robs (primarily higher-mass) satellites of their orbital energy, one might expect satellites to shrink monotonically in orbital radius over time, until the satellite merges/disrupts. Many studies implement idealized simulations (both with and without baryonic physics) that incorporate the effects of dynamical friction, mass loss, and ram-pressure stripping

(for example Weinberg, 1986; Taylor & Babul, 2001; Peñarrubia et al., 2002; Peñarrubia & Benson, 2005; Amorisco, 2017; Jiang et al., 2021), but testing this assumption in a cosmological setting is imperative.

Additionally, because the LMC has satellites of its own (for example D'Onghia & Lake, 2008; Deason et al., 2015; Kallivayalil et al., 2018), many studies have investigated satellite spatial distributions and their dynamics with an additional LMC-like contribution to the host potential to test the dynamical effects of a massive satellite on nearby lower-mass galaxies (for example Garavito-Camargo et al., 2019; Patel et al., 2020; Samuel et al., 2021; Vasiliev et al., 2021; D'Souza & Bell, 2022; Pace et al., 2022).

In this paper, we examine the orbital histories of satellite galaxies using baryonic simulations that match general observed properties of satellites of MW-mass galaxies. Specifically, we use the FIRE-2 cosmological zoom-in simulations, which form realistic MW-mass galaxies (for example Garrison-Kimmel et al., 2018; Hopkins et al., 2018; Sanderson et al., 2020; Bellardini et al., 2021) and populations of satellite galaxies around them (Wetzel et al., 2016; Garrison-Kimmel et al., 2019b; Samuel et al., 2020, 2022b). The two main topics that we explore are: (1) The relation of orbital properties of satellite galaxies at $z = 0$ to their orbital histories, including lookback times of infall, distances from the MW-mass host, and stellar masses. These relations not only help characterize the orbits of satellites, but their orbital histories also provide insight, or caution, into approximating history-based properties, such as infall time or pericentre information, based on their present-day properties, such as distance and stellar mass. (2) Testing a common expectation that the orbits of satellite galaxies shrink over time, that is, that a satellite's most recent pericentric distance is the minimum that it has experienced.

In Santistevan et al., in prep, we will compare directly the orbits of satellites in cosmological simulations of MW-mass galaxies to orbits from integration in a static, idealized MW-mass halo (see also Vasiliev et al., 2021; D'Souza & Bell, 2022, and references therein).

4.3. Methods

TABLE 4.1. Properties at $z = 0$ of the 13 MW/M31-mass galaxies in the FIRE-2 simulation suite that we analyze, ordered by decreasing stellar mass. Simulations with 'm12' names are isolated galaxies from the Latte suite, while the others are from the 'ELVIS on FIRE' suite of LG-like paired hosts. Columns: host name; $M_{\rm star,90}$ is the host's stellar mass within $R_{\rm star,90}$, the disk radius enclosing 90 per cent of the stellar mass within 20 kpc; $M_{\rm 200m}$ is the halo total mass; $R_{\rm 200m}$ is the halo radius; $N_{\rm satellite}$ is the number of satellite galaxies at $z = 0$ with $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$ that ever orbited within $R_{\rm 200m}$; $M^{\rm max}_{\rm sat,star}$ is the stellar mass of the most massive satellite at $z = 0$; and $M^{\rm max}_{\rm sat,halo,peak}$ is the peak halo mass of the most massive satellite. In Remus and Juliet, the satellite with the largest stellar mass is not the same as the satellite with the largest subhalo mass.

Name	$M_{\rm star,90}$ $[10^{10}\,{\rm M}_\odot]$	$M_{\rm 200m}$ $[10^{12}\,{\rm M}_\odot]$	$R_{\rm 200m}$ [kpc]	$N_{\rm satellite}$	$M^{\rm sat}_{\rm star,max}$ $[10^{8}\,{\rm M}_\odot]$	$M^{\rm sat}_{\rm halo,peak,max}$ $[10^{10}\,{\rm M}_\odot]$	Reference
m12m	10.0	1.6	371	44	4.68	3.78	A
Romulus	8.0	2.1	406	52	2.34	2.68	B
m12b	7.3	1.4	358	32	0.58	1.21	C
m12f	6.9	1.7	380	43	1.61	2.01	D
Thelma	6.3	1.4	358	33	0.33	1.78	C
Romeo	5.9	1.3	341	33	1.92	3.34	C
m12i	5.5	1.2	336	26	1.24	2.40	E
m12c	5.1	1.4	351	40	15.1	16.7	C
m12w	4.8	1.1	319	38	7.79	4.88	F
Remus	4.0	1.2	339	34	0.50*	2.07*	B
Juliet	3.3	1.1	321	38	2.51*	3.16*	C
Louise	2.3	1.2	333	34	2.50	3.71	C
m12r	1.5	1.1	321	27	28.4	13.7	F

Note: Simulation introduced in: A: Hopkins et al. (2018), B: Garrison-Kimmel et al. (2019a), C: Garrison-Kimmel et al. (2019b), D: Garrison-Kimmel et al. (2017b), E: Wetzel et al. (2016), F: Samuel et al. (2020).

4.3.1. FIRE-2 Simulations. We use the cosmological zoom-in baryonic simulations of both isolated MW/M31-mass galaxies and LG-like pairs from the Feedback In Realistic Environments (FIRE) project[1] (Hopkins et al., 2018). We ran these simulations using the hydrodynamic plus N-body code GIZMO (Hopkins, 2015), with the mesh-free finite-mass (MFM) hydrodynamics method (Hopkins, 2015). We used the FIRE-2 physics model (Hopkins et al., 2018) that includes

[1]See the FIRE project web site: http://fire.northwestern.edu

several radiative heating and cooling processes such as Compton scattering, Bremsstrahlung emission, photoionization and recombination, photoelectric, metal-line, molecular, fine-structure, dust-collisional, and cosmic-ray heating across temperatures $10 - 10^{10}$ K, including the spatially uniform and redshift-dependent cosmic ultraviolet (UV) background from Faucher-Giguère et al. (2009), for which HI reionization occurs at $z_{\rm reion} \approx 10$. Star formation occurs in gas that is self-gravitating, Jeans unstable, molecular (following Krumholz & Gnedin, 2011), and dense ($n_{\rm H} > 1000$ cm^{-3}). Star particles represent single stellar populations, assuming a Kroupa (2001) initial mass function, and they evolve along stellar population models from STARBURST99 v7.0 (Leitherer et al., 1999), inheriting masses and elemental abundances from their progenitor gas cells. FIRE-2 simulations also include the following stellar feedback processes: core-collapse and Ia supernovae, stellar winds, and radiation pressure.

We used the code MUSIC (Hahn & Abel, 2011) to generate the cosmological zoom-in initial conditions at $z \approx 99$ within periodic cosmological boxes of length $70.4 - 172$ Mpc, sufficiently large to avoid unrealistic periodic gravity effects on individual MW-mass halos. For each simulation, we save 600 snapshots with $20 - 25$ Myr spacing down to $z = 0$, assuming a flat ΛCDM cosmology. Consistent with Planck Collaboration et al. (2018), we used cosmological parameters in the following ranges: $h = 0.68 - 0.71$, $\sigma_8 = 0.801 - 0.82$, $n_{\rm s} = 0.961 - 0.97$, $\Omega_\Lambda = 0.69 - 0.734$, $\Omega_{\rm m} = 0.266 - 0.31$, and $\Omega_{\rm b} = 0.0449 - 0.048$.

Our galaxy sample consists of the 12 MW/M31-mass galaxies in Santistevan et al. (2020), as well as one additional galaxy, 'm12r' first introduced in Samuel et al. (2020). These are from the Latte suite of isolated MW/M31-mass galaxies introduced in Wetzel et al. (2016) and the 'ELVIS on FIRE' suite of LG-like MW+M31 pairs, introduced in Garrison-Kimmel et al. (2019b). Table 4.1 lists their stellar mass, $M_{\rm star,90}$, halo mass, $M_{\rm 200m}$, and radius $R_{\rm 200m}$ at $z = 0$, where both are defined at 200 times the matter density at $z = 0$. The Latte suite consists of halos with $M_{\rm 200m} = 1 - 2 \times 10^{12}\,{\rm M}_\odot$ at $z = 0$ with no other similar-mass halos within $5 \times R_{\rm 200m}$. We also chose m12r and m12w to have LMC-mass satellite analogs near $z \approx 0$ (Samuel et al., 2020).

The initial masses of star particles and gas cells is $7100\,M_\odot$, but the average star particle mass at $z = 0$ is $\approx 5000\,M_\odot$ from stellar mass loss. Within the zoom-in region, the mass of DM particles is $3.5\times10^4\,M_\odot$. The gravitational softening lengths are 4 and 40 pc (Plummer equivalent), co-moving at $z > 9$ and physical thereafter, for star and DM particles, respectively. Gas cells use adaptive force softening, equal to their hydrodynamic smoothing, down to 1 pc.

Each pair of halos in the 'ELVIS on FIRE' suite of LG-like pairs was chosen based on their individual mass ($M_{200m} = 1 - 3 \times 10^{12}\,M_\odot$) and combined masses (total LG mass between $2 - 5 \times 10^{12}\,M_\odot$), as well as their current separation ($600 - 1000$ kpc) and radial velocities at $z = 0$ ($v_{rad} < 0$), and isolated environment (no other massive halos within 2.8 Mpc of either host center). The mass resolution is $\approx 2\times$ better in the 'ELVIS on FIRE' suite, with initial baryonic particle masses $3500 - 4000\,M_\odot$. For all results in this paper, we investigated possible differences between the isolated and LG-like MW-mass galaxies, which also partially tests resolution convergence, and we find negligible differences between the two samples.

These 13 host galaxies reflect formation histories of general MW/M31-mass (or LG-like) galaxies within our selection criteria and exhibit observational properties broadly similar to the MW and M31, including: realistic stellar halos (Bonaca et al., 2017; Sanderson et al., 2018b), dynamics of metal-poor stars from early galaxy mergers (Santistevan et al., 2021), satellite galaxy stellar masses and internal velocity dispersions (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a), radial and 3-D spatial distributions (Samuel et al., 2020, 2021), and star-formation histories and quiescent fractions (Garrison-Kimmel et al., 2019a; Samuel et al., 2022b).

4.3.2. Halo/Galaxy Catalogs and Merger Trees. We use the ROCKSTAR 6-D halo finder (Behroozi et al., 2013a) to generate (sub)halo catalogs using only DM particles at each of the 600 snapshots, and CONSISTENT-TREES (Behroozi et al., 2013b) to generate merger trees. As a consequence of the large zoom-in volume for each host, there is no low-resolution DM particle contamination in any of the (sub)halos that we analyze.

Samuel et al. (2020) describes our star particle assignment to (sub)halos in post-processing; we briefly review it here. We first select star particles within $d < 0.8R_{\rm halo}$ (out to a maximum distance of 30 kpc) with velocities $v < 2V_{\rm circ,max}$ of the (sub)halo's center-of-mass (COM) velocity. We then keep only the star particles within $d < 1.5R_{\rm star,90}$ (the radius enclosing 90 per cent of the stellar mass) of the (then) current member stellar population's COM and halo center position. We further kinematically select the star particles with velocities $v < 2\sigma_{\rm vel,star}$ (the velocity dispersion of current member star particles) of the COM velocity of member star particles. Finally, we iterate on these spatial and kinematic criteria, which guarantees that the COM of the galaxy and (sub)halo are consistent with one another, until the (sub)halo's stellar mass converges to within 1 per cent.

We use two publicly available analysis packages: HaloAnalysis[2] (Wetzel & Garrison-Kimmel, 2020a) for assigning star particles to halos and for reading and analyzing halo catalogs/trees, and GizmoAnalysis[3] (Wetzel & Garrison-Kimmel, 2020b) for reading and analyzing particles from Gizmo snapshots.

4.3.3. Selection of Satellites. We include all luminous satellite galaxies at $z = 0$ with $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$ that have crossed within their MW-mass host halo's $R_{\rm 200m}(z)$. This lower limit on stellar mass corresponds to ≈ 6 star particles, the limit for reasonably resolving the total stellar mass (Hopkins et al., 2018). Our sample includes 'splashback' satellites that are currently beyond the host's $R_{\rm 200m}$, which are typically still gravitationally bound to the host but simply near apocentre (for example Wetzel et al., 2014). As Table 4.1 shows, the number of surviving luminous satellites at $z = 0$, including this splashback population, per host ranges from 26-52, and our sample totals 473 satellites. Both Garrison-Kimmel et al. (2014) and Wetzel et al. (2015) showed that in the ELVIS DMO simulation suite the average number of subhalos that would typically host galaxies with $M_{\rm star} \gtrsim 10^5\,{\rm M}_\odot$ is $\sim 31 - 45$. However, because stellar feedback and baryonic physics can affect galaxy formation, Wetzel et al. (2016) and Garrison-Kimmel et al. (2019b) showed that the number of satellites above this mass range decreases to $\sim 13 - 15$. More recently, Samuel et al.

[2] https://bitbucket.org/awetzel/halo_analysis
[3] https://bitbucket.org/awetzel/gizmo_analysis

(2020) showed that the radial distributions of these satellites are consistent with the MW and M31 out to 300 kpc. The MW and M31 each have 13 and 27 satellites, respectively, and the MW-mass hosts in our simulations bracket these values with 11 to 27 satellites. Unless otherwise stated, in our analysis we refer to luminous satellites, i.e. satellites containing stars, as simply 'satellites'.

In computing host-averaged results below, to avoid biasing our results to the hosts with larger satellite populations, we oversample the satellites so that each host contributes a nearly equal fraction of satellites to the total. Specifically, we multiply the number of satellites in the MW-mass host with the largest population (Romulus, with 52 satellites) by 10 which results in an oversampled population of 520. Then, for each of the other MW-mass hosts, we divide 520 by the number of their satellites and obtain the nearest integer multiplicative factor, m, that we apply to each host's satellite population, $N_{\rm sat}$ (see Table 4.1), so that each host contains $\approx 500 - 530$ satellites, or that their satellite populations are within 5 per cent of one another. Thus, when plotting properties, such as pericentre distances, we count each satellite in a given host m times for each property in the figures.

Figure 4.1 caption: Stellar mass, $M_{\rm star}$, at $z = 0$ versus peak halo mass, $M_{\rm halo,peak}$, for all 473 satellites with $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$ across all 13 host galaxies. The solid blue line shows the median, and the dark and light shaded regions show the 68th percentile and full distribution, respectively. We compare this with the stellar-halo mass relation of non-satellite low-mass galaxies which never fell into the MW-mass host, and currently orbit beyond 1 Mpc at $z = 0$ (blue dot-dashed). The dotted grey line indicates the minimum $M_{\rm star}$ in our sample. For comparison, we also show extrapolations of the stellar-halo mass relations from Moster et al. (2013) (red), Garrison-Kimmel et al. (2017a) (green), and Behroozi et al. (2020) (black), which broadly agree with our sample at $M_{\rm halo,peak} \gtrsim 10^9\,{\rm M}_\odot$. Additionally, we show the 68th percentile and full distribution of values from higher resolution, *isolated* low-mass galaxies from Fitts et al. (2017) in the dark and light shaded pink regions, respectively, and the median across their sample in the horizontal pink line. We finally show 4 higher resolution, isolated low-mass galaxies from Wheeler et al. (2019) as stars,

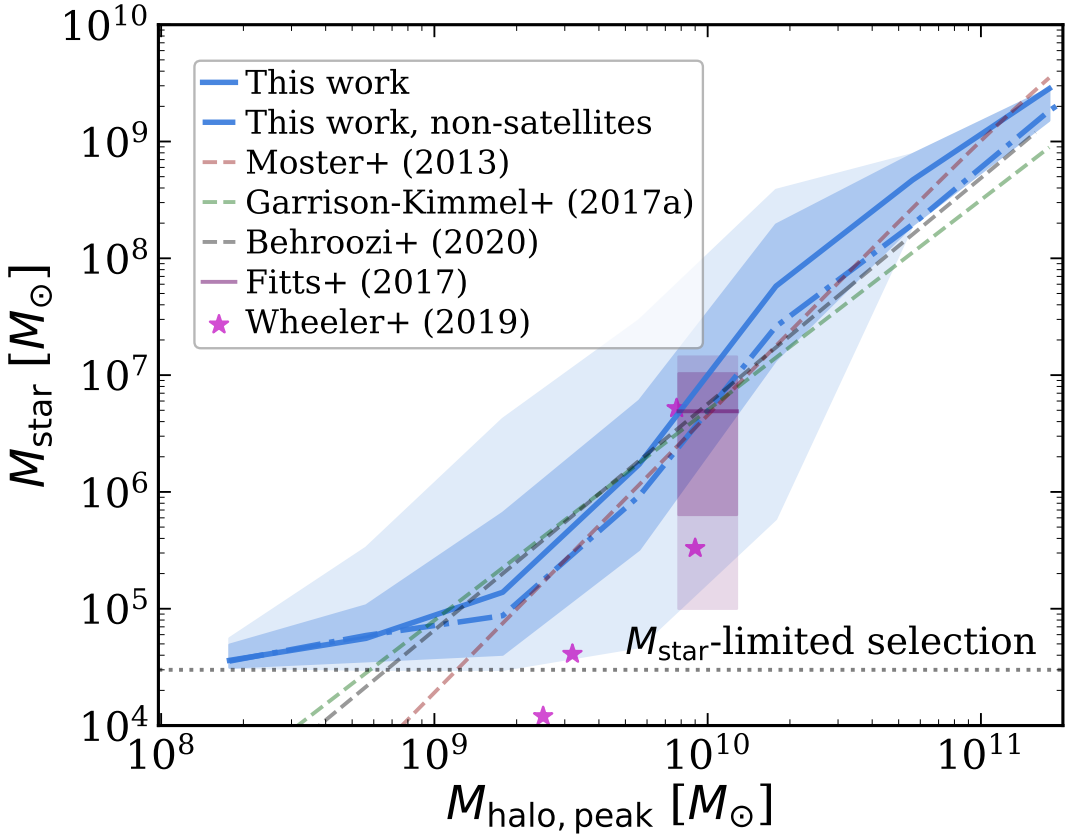

FIGURE 4.1. See the figure caption in the text.

however, the galaxy with the smallest $M_{\rm star}$ in this sample is beyond our resolution limit. At all halo masses, the 68th percentile in satellite $M_{\rm star}$ is 0.5 – 1 dex, with a smaller range at the extreme masses from lower statistics. The relation flattens at $M_{\rm star} \lesssim 10^5\,{\rm M}_\odot$ simply because of our stellar mass limit; **therefore, our fiducial sample is complete in $M_{\rm star}$ to $> 3 \times 10^4\,{\rm M}_\odot$, while we are complete in $M_{\rm halo,peak}$ to $\gtrsim 3 \times 10^9\,{\rm M}_\odot$** (see Appendix A.1 for results for selecting satellites via $M_{\rm halo,peak}$).

Figure 4.1 shows the relation between stellar mass and halo mass (SMHM) for our satellites. We show stellar mass *at* $z = 0$ versus *peak* (sub)halo mass throughout its history. We include the median SMHM relation of non-satellites in the same simulations, which we define as low-mass galaxies that never crossed the virial radius of the MW-mass host, and that currently orbit beyond

1 Mpc at $z = 0$, in the blue dot-dashed line. The formation histories of non-satellites differ from satellites because they form in less dense regions of the Universe. This, along with the UV heating of gas in isolated low-mass galaxies may explain their slightly smaller $M_{\rm star}$, however, a deeper investigation is outside of the scope of this paper. The grey dotted line at $M_{\rm star} = 3 \times 10^4\,{\rm M}_\odot$ shows our lower limit in stellar mass.

To compare to other SMHM relations from the literature, we also show extrapolations from Moster et al. (2013); Garrison-Kimmel et al. (2017a); Behroozi et al. (2020) in the dashed red, green, and black lines, respectively. We also include the median $M_{\rm star}$, 68th percentile, and full distribution of values from Fitts et al. (2017) of higher resolution (initial baryon masses of $500\,{\rm M}_\odot$), isolated low-mass galaxies, in the pink horizontal line, and dark and light pink shaded regions, respectively. Finally, we include 4 higher resolution, isolated low-mass galaxies from Wheeler et al. (2019) (named m09, m10q, m10v, and m10vB in their work), with initial baryon masses of $30\,{\rm M}_\odot$ (magenta stars). The SMHM relation in our sample broadly agrees with these (extrapolated) semi-empirical estimates and values of isolated low-mass galaxies; for a more detailed discussion about the SMHM relation in FIRE-2, see Hopkins et al. (2018); Wheeler et al. (2019). The low-mass end of our SMHM relation in Figure 4.1 flattens at $M_{\rm star} \lesssim 10^5\,{\rm M}_\odot$, purely because of our stellar mass selection of $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$. We note that the galaxy with the smallest $M_{\rm star}$ from Wheeler et al. (2019) is beyond the resolution limit in our sample, and the second smallest $M_{\rm star}$ galaxy would be only marginally resolved. However, with better resolution, the SMHM relation in Figure 4.1, would likely follow a similar trend as the extrapolations and it is likely that the lowest $M_{\rm star}$ isolated galaxy would lie in the lower end of the full distribution scatter. Thus, while our sample is complete in stellar mass, we are complete in halo mass for $M_{\rm halo,peak} \gtrsim 3 \times 10^9\,{\rm M}_\odot$. We find only minor differences in our results if selecting satellites via $M_{\rm halo,peak}$ instead (see Appendix A.1).

We checked the SMHM relation in Figure 4.1 instead using the *peak* stellar mass throughout a galaxy's history, $M_{\rm star,peak}$. The two relations are similar, but the SMHM relation with $M_{\rm star,peak}$ has a $\approx 5\%$ higher normalization, on average, because of stellar mass loss after infall.

4.3.4. Numerical Disruption. Many previous studies (for example van den Bosch et al., 2018; van den Bosch & Ogiya, 2018; Webb & Bovy, 2020) noted that without proper mass and spatial resolution, satellite subhalos can suffer from artificial numerical disruption too quickly. Thus, sufficient resolution and implementation of the relevant physics is necessary to model accurately the evolution of satellite galaxies. As a partial test of this, we investigated differences between the isolated and LG-like satellite populations, which have a $\approx 2\times$ resolution difference, and we saw no strong differences in our results. Samuel et al. (2020) also tested resolution convergence of the satellite radial profiles using the FIRE-2 simulations with dark matter particle masses of $m_{\rm dm} = 3.5 \times 10^4\,{\rm M}_\odot$ and $m_{\rm dm} = 2.8 \times 10^5\,{\rm M}_\odot$ and found generally consistent results between the two, because there are enough dark matter particles ($\gtrsim 2 \times 10^4$) in the lowest-mass luminous subhalos ($M_{\rm halo,peak} \gtrsim 10^8\,{\rm M}_\odot$) to prevent numerical disruption, and many more than that in our typical (more massive) luminous subhalos, thus satisfying criteria such as in van den Bosch & Ogiya (2018). Also, our DM particle mass resolution is $m_{\rm dm} = 2 - 3.5 \times 10^4\,{\rm M}_\odot$ and DM force softening length is 40 pc, significantly better than previous work like Wetzel et al. (2015), who used the ELVIS DMO simulations with $m_{\rm dm} = 1.9 \times 10^5\,{\rm M}_\odot$ and 140 pc but found results broadly consistent with ours. Nonetheless, any simulations like ours necessarily operate at finite resolution, which inevitably leads to some degree of numerical disruption, which any reader should bear in mind.

4.3.5. Calculating pericentre. We calculate pericentres by tracking the main progenitor back in time using the merger trees, which store each satellite's galactocentric distance at each of the 600 snapshots. We first ensure that the satellite is within the MW-mass host halo at a given snapshot. Then, we check if the galactocentric distance reaches a minimum within ±20 snapshots, corresponding to a time window of ≈ 1 Gyr. Given the ≈ 25 Myr time spacing between snapshots, we fit a cubic spline to the distance and time arrays across this interval. We then find the minimum in the spline-interpolated distance(s), and record the corresponding spline-interpolated time.

We checked how our results differ if varying the window to ± 4, 8, and 10 snapshots by visually inspecting each satellite's orbit history and conclude that a window of ±20 snapshots reduces nearly all 'false' pericentres, that is, instances in which the criteria above are met because of numerical noise in the orbit or a short-lived perturbation. Because the center of mass of the MW-mass galaxy does not perfectly coincide with the center of mass of its DM host halo, we also checked how our results vary between using distances with respect to the center of the galaxy versus the halo: we find no meaningful difference. We additionally checked how our results vary when using distances with respect to the center of the satellite galaxy versus the center of the satellite (sub)halo, again finding no significant differences.

4.3.6. Calculating Gravitational Potential and Total Energy. We also explore trends with a satellite's specific orbital total energy, E at $z = 0$, defined as the sum of the kinetic and potential energies. Our simulations store the value of the gravitational potential for all particles at each snapshot, so to calculate the potential for each satellite at $z = 0$, we select all star, gas, and DM particles within ±5 kpc of the satellite (sub)halo's virial radius, to limit biasing from the satellite's self-potential, and we compute the mean potential across these particles. Given that some satellites are in LG-like environments, we normalize E at the MW-mass halo radius, such that $E(d = R_{200\mathrm{m}}) = 0$, that is, the sum of the host potential at $R_{200\mathrm{m}}$ and the kinetic energy of a circular orbit at $R_{200\mathrm{m}}$ is 0.

4.4. Results

Throughout the paper, we present results for all satellite galaxies at $z = 0$, based on their stellar mass ($M_{\mathrm{star}} > 3 \times 10^4\,\mathrm{M}_\odot$), across all of our MW-mass hosts. Although Santistevan et al. (2020) noted that MW-mass halos/galaxies in LG-like pairs formed ~ 1.6 Gyr earlier than those that are isolated, we compared satellites based on isolated versus LG-like environment and find negligible differences in any properties that we investigate. This agrees with the lack of dependence in Wetzel et al. (2015), who investigated the infall histories of satellite subhalos in the ELVIS dark

matter-only simulations. Appendix A.1 examines how our results change by selecting satellites by their peak halo mass. In summary we find qualitatively similar results for selecting via stellar mass, but given the scatter in the SMHM relation, the trends with halo mass are smoother and any features are sharper. Although the dark matter (sub)halo mass is more dynamically relevant to the orbits of satellite galaxies, we present our results with a stellar mass selected sample, because it is observationally easier to measure than halo mass. Finally, Appendix A.2 compares our results for baryonic simulations against DMO simulations of the same systems. In summary, the lack of a MW-mass galaxy in the DMO simulations allows satellites to survive longer, orbit closer to the center of the halo, and complete more orbits.

We examine trends guided by which orbital properties are relevant to different phenomena. As we will show, specific angular momentum and specific total energy provides insight into when a satellite fell into the MW-mass halo. We also explore trends with satellite mass, in part to understand where dynamical friction becomes important: from Equation 4.1, for the MW-mass halos with $M_{200\mathrm{m}}(z = 0) \approx 10^{12}\,\mathrm{M}_\odot$ (and lower $M_{200\mathrm{m}}$ at earlier times), we expect dynamical friction to significantly affect satellites with $M_{200\mathrm{m}} \gtrsim 3 \times 10^{10}\,\mathrm{M}_\odot$, or $M_{\mathrm{star}} \gtrsim 10^{8}\,\mathrm{M}_\odot$. We also focus on infall times, to understand how long satellites have been orbiting in the host halo environment, and we explore the incidence of pre-processing in a lower-mass group prior to MW-mass infall. We also examine properties of orbital pericentre, given that satellites typically feel the strongest gravitational tidal force and the strongest ram pressure at pericentre.

In this paper, we present trends for the simulated satellite populations only, however, in the future we plan to investigate differences in the simulations and results obtained from idealized orbit modeling methods. Ultimately we will provide a framework to derive similar orbital properties from satellites in the MW and M31 using the satellite populations in the simulations, and compare the results. Thus, we leave direct observational comparisons for future work.

4.4.1. Orbital properties today. *Figure 4.2 caption:* Instantaneous orbital dynamics of satellite galaxies at $z = 0$ versus their lookback time of infall into the MW-mass halo (left), distance

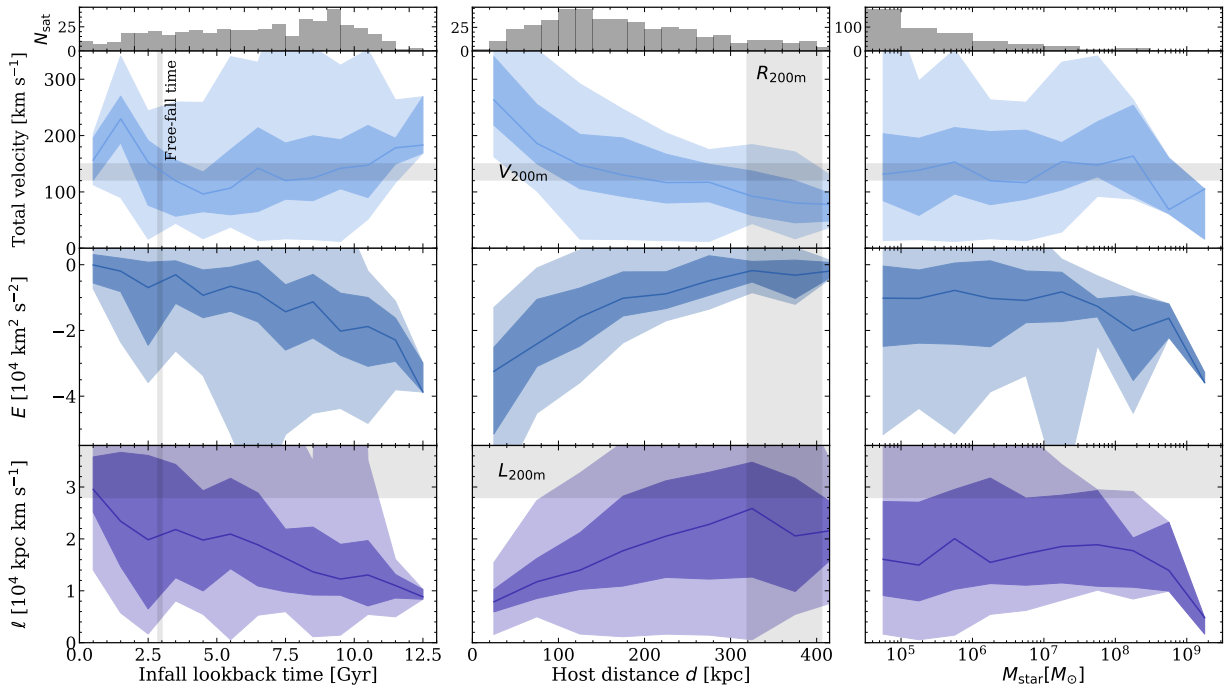

FIGURE 4.2. See the figure caption in the text.

from MW-mass host, d (middle), and stellar mass, $M_{\rm star}$ (right). The solid lines show the median for all satellites across our 13 MW-mass hosts, and the dark and light shaded regions show the 68th percentile and full distribution, respectively. At the top of each column, we show the histogram of these properties for the sample. We show common reference points to these properties for typical host halos within our mass range in the horizontal and vertical gray shaded bands in some of these panels, such as the free-fall time, and virial radii, velocities, and angular momenta; see text for details. **Top row**: Orbital total velocity. The median velocity for satellites that fell in $> 4\,$Gyr ago increases with earlier infall time from $100 - 200\,{\rm km\,s^{-1}}$ (top left). However, more recently infalling satellites show a peak near $\approx 1.5\,$Gyr of $\approx 230\,{\rm km\,s^{-1}}$, highlighting satellites that are nearing their first pericentre, before decreasing again to $\approx 100\,{\rm km\,s^{-1}}$ for satellites that only recently fell in. Velocity also decreases with host distance and is relatively flat with $M_{\rm star}$ at roughly $135\,{\rm km\,s^{-1}}$. **Middle row**: Specific orbital total energy, E. We normalize E at the host's $R_{\rm 200m}$ so $E(d = R_{\rm 200m}) = 0$. E decreases for earlier-infalling satellites, with median values

ranging from 0 to $-3.8 \times 10^4\,\mathrm{km^2\,s^{-2}}$. Satellites that fell in earlier orbit deeper in the host halo gravitational potential and therefore are more bound (middle left). Satellites with $M_{\rm star} \gtrsim 10^8\,\mathrm{M_\odot}$ are more bound, because of the stronger dynamical friction they experience, and satellites closer to the host are more bound, because they orbit deeper in the gravitational potential (middle right). **Bottom row**: Orbital specific angular momentum, ℓ. Specific angular momentum decreases for earlier-infalling satellites, from ≈ 3 to $0.9 \times 10^4\,\mathrm{kpc\,km\,s^{-1}}$ (bottom left), because satellites that fell in earlier fell in (and orbit) at smaller distances (see Figure 4.5 and bottom left panel of Figure 4.6). ℓ necessarily increases with d (bottom middle) and depends only weakly on satellite $M_{\rm star}$ (bottom right).

We first investigate the instantaneous orbital properties of satellites at $z = 0$, including approximate integrals of motion like angular momentum and energy. Figure 4.2 shows total velocity, specific orbital total energy, E, and specific angular momentum, ℓ, as a function of lookback time of infall into the MW-mass halo, $t^{\rm lb}_{\rm infall,MW}$, galactocentric distance, d, and stellar mass, $M_{\rm star}$. The top of each column shows distributions of $t^{\rm lb}_{\rm infall,MW}$, d, and $M_{\rm star}$. In particular, the distribution of $t^{\rm lb}_{\rm infall,MW}$ is relatively flat, with a modest peak $\approx 9\,\mathrm{Gyr}$ ago. For reference, all panels versus $t^{\rm lb}_{\rm infall,MW}$ (left) include a vertical shaded region to represent the free-fall time at $R_{\rm 200m}$, $t_{\rm ff} = \sqrt{3\pi/(32 G \rho_{\rm 200m})}$, where $t_{\rm ff} = 2.8 - 3\,\mathrm{Gyr}$ across our hosts. In all panels versus d (middle), the vertical shaded region shows the range of $R_{\rm 200m}$ for the MW-mass halos, as Table 4.1 lists. Similarly, all panels versus total velocity and ℓ show horizontal shaded bands that represent the range of $V_{\rm 200m}$ and $L_{\rm 200m}$ across hosts, where $V_{\rm 200m} = \sqrt{G M_{\rm 200m}/R_{\rm 200m}}$ is the velocity of a circular orbit at the virial radius, and $L_{\rm 200m} = V_{\rm 200m} \times R_{\rm 200m}$ is the specific angular momentum of that orbit. Because the satellites fell in at different times and orbit at various distances with unique stellar masses, we do not expect them to have values equal to $V_{\rm 200m}$ or $L_{\rm 200m}$, so we provide these shaded regions as a reference only.

4.4.1.1. *Total velocity.* We first present trends in the total velocity, corresponding to the top row in Figure 4.2. Considering the trends in total velocity with infall time (top left), satellites that fell in $< 1\,\mathrm{Gyr}$ ago have not yet experienced a pericentre, so they show an increase in total velocity

with time since infall from 0.5 – 1.5 Gyr ago, because these satellites are near pericentre. The total velocity then decreases with increasing time since infall, because those satellites are now near apocentre. We see a similar, but weaker, peak in total velocity ~ 6 Gyr ago, from a marginally phase-coherent population near pericentre, but after a few orbits, satellites become out of phase. Satellites that fell in $\lesssim 3$ Gyr ago typically have total velocities of $150 - 250\,\mathrm{km\,s^{-1}}$, while earlier infalling satellites are only orbiting at $100 - 185\,\mathrm{km\,s^{-1}}$. Comparing these infall times to $t_{\rm ff}$ in the vertical shaded bands, satellites with $t^{\rm lb}_{\rm infall,MW} < t_{\rm ff}$ often have larger velocities, and again are likely to be on first infall, because they have not had enough time to reach pericentre. For reference, we also compare the satellite total velocities to the host's virial velocity, $V_{\rm 200m}$ (horizontal gray shaded band). Satellites that fell in $t^{\rm lb}_{\rm infall,MW} \approx 3 - 11$ Gyr ago typically have comparable total velocities to the virial velocity, but the full scatter extends to both much larger and smaller values.

Next, the median velocity decreases with increasing d (top middle), from as high as $260\,\mathrm{km\,s^{-1}}$ for the closest satellites to $80\,\mathrm{km\,s^{-1}}$ for satellites near $R_{\rm 200m}$. The shaded region of the 68th percentile follows the median trend and the width is roughly constant at $\approx 95\,\mathrm{km\,s^{-1}}$ across all distances. At $R_{\rm 200m}$, the total velocities of satellites are typically lower than $V_{\rm 200m}$ both because they are not on perfectly circular orbits and because the population at large d is likely biased to lower values from the splashback population which typically have negligible velocities.

The median is nearly constant with $M_{\rm star}$ (top right), ranging from $120 - 160\,\mathrm{km\,s^{-1}}$ for $M_{\rm star} \approx 10^{4.75-8.25}\,\mathrm{M_\odot}$, which decreases to $70 - 100\,\mathrm{km\,s^{-1}}$ at higher mass, likely because sufficiently massive satellites experience significant dynamical friction that slows their orbit. Across all stellar masses, the average total velocity is $\approx 135\,\mathrm{km\,s^{-1}}$ and the typical range of the 68th percentile is $90 - 205\,\mathrm{km\,s^{-1}}$. The median for all satellites with $M_{\rm star} \lesssim 10^{8.5}\,\mathrm{M_\odot}$ show consistent values with $V_{\rm 200m}$.

4.4.1.2. *Specific total energy.* Next, the middle row in Figure 4.2 shows trends in the specific orbital energy, E. Given that some satellites are in LG-like environments, which complicates computing the specific total energy beyond $R_{\rm 200m}$, we normalize E at the MW-mass halo's $R_{\rm 200m}$

so $E(d = R_{200\mathrm{m}}) = 0$. Thus, satellites with $E > 0$ are the splashback population with apocentres beyond $R_{200\mathrm{m}}$ and are essentially all bound to the host halo.

The middle left panel shows that earlier infalling satellites are on more bound orbits, with the median E decreasing from ≈ 0 to $-3.9 \times 10^4\,\mathrm{km^2\,s^{-2}}$. This reflects the growth of the MW-mass halo over time. E increases with d from the host, so satellites are more bound at smaller distances. Thus, the median E is not constant with d, but this does not imply that specific energy is not being conserved over an orbit. As we explore below, $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$ correlates with d, such that satellites at smaller d fell in earlier (though with large scatter; see Figure 4.5). This is largely because they fell in when the host $R_{200\mathrm{m}}$ was smaller, so they necessarily orbit at smaller d. This then leads to the correlation of E with d across the population at $z = 0$ (Figure 4.2, center panel).

Similar to the trends in total velocity, E does not strongly depend on M_{star}, except at $M_{\mathrm{star}} \gtrsim 10^8\,\mathrm{M_\odot}$, where satellites experience significant dynamical friction, causing their orbits to become more bound, despite the fact that higher-mass satellites fell in more recently, as we show below.

4.4.1.3. *Specific angular momentum.* Last, we present trends in specific angular momentum in the bottom row of Figure 4.2 The median specific angular momentum, ℓ (bottom), decreases across time since infall from ≈ 3 to $1 \times 10^4\,\mathrm{kpc\,km\,s^{-1}}$ (bottom left panel). Between $t^{\mathrm{lb}}_{\mathrm{infall,MW}} \approx 2-6\,\mathrm{Gyr}$, ℓ is nearly constant at $\approx 2 \times 10^4\,\mathrm{kpc\,km\,s^{-1}}$, which as we will show in Figure 4.6 corresponds to satellites that completed $1-2$ pericentres. The median ℓ is much higher for satellites that are to the left of the t_{ff} band, consistent with the higher velocities in these satellites. Similar to the top row, we define the reference host halo angular momentum, $L_{200\mathrm{m}} = V_{200\mathrm{m}} \times R_{200\mathrm{m}}$, and we show the range of values across hosts in the gray horizontal band. As expected, essentially all satellites have $\ell < L_{200\mathrm{m}}$, and most have ℓ significantly lower.

Generally, ℓ and its scatter increase with increasing d as expected given that $\ell = v_{\mathrm{tan}} d$, where v_{tan} is the satellite's tangential velocity. Because the MW-mass halo grew over time, satellites fell into their MW-mass halo on larger orbits at later times, which explains the increasing ℓ with more recent infall times.

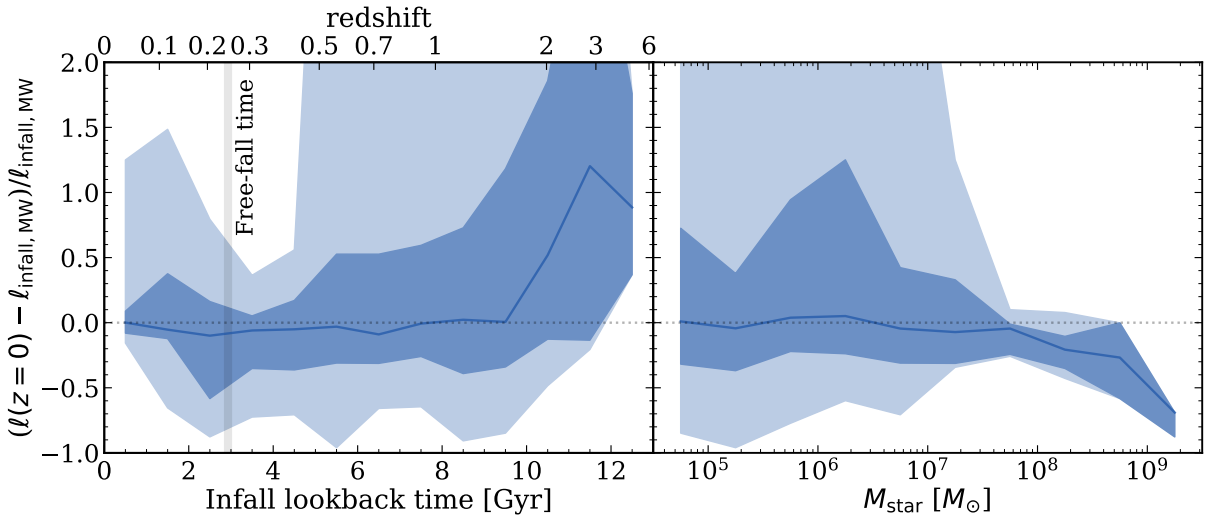

FIGURE 4.3. The fractional difference in specific angular momentum since infall into the MW-mass halo, $(\ell_{\rm now} - \ell_{\rm infall,MW})/\ell_{\rm infall,MW}$. Lines show the median, and the dark and light shaded regions represent the 68th percentile and full distributions, respectively. Similar to Figure 4.3, the vertical gray band shows the free-fall time at the host virial radius. **Left**: The median fractional difference in ℓ is within ≈ 10 per cent for satellites that fell in $t^{\rm lb}_{\rm infall,MW} \lesssim 10\,$Gyr ago, and the typical scatter about this median is ≈ 40 percent. However, earlier-infalling satellites *increased* ℓ, on average, up to twice their value at infall. **Right**: Lower-mass satellites show small typical changes in ℓ since infall, though again with a typical scatter of $\approx 30-50$ per cent. At $M_{\rm star} \gtrsim 10^8$, satellites decreased in ℓ, likely from stronger dynamical friction. *Thus, specific angular momentum of satellites is reasonably conserved on average, but with large (≈ 40 percent) scatter, for satellites with $M_{\rm star} \lesssim 10^8$ that fell in $\lesssim 10\,$Gyr ago. Higher-mass satellites, or those that fell in earlier, can experience significant changes in ℓ.*

The median ℓ is nearly constant with stellar mass at $M_{\rm star} < 10^{8.25}\,{\rm M}_\odot$, but as with velocity, it decreases for higher-mass satellites. Although we find little dependence of median d with $M_{\rm star}$ (not shown), satellites with $M_{\rm star} > 10^8\,{\rm M}_\odot$ today exist at $\lesssim 250\,$kpc, whereas there are lower mass satellites out to $\approx 800\,$kpc. Thus, because dynamical friction likely drove higher-mass satellites to smaller velocities and distances, they have smaller orbital lifetimes and smaller ℓ today. Across the full mass range, the mean of the median is $1.6 \times 10^4\,{\rm kpc\,km\,s^{-1}}$, and the mean range of the 68th percentile is $0.9 - 2.5 \times 10^4\,{\rm kpc\,km\,s^{-1}}$.

To investigate how ℓ changed over time, we measure the fractional difference in ℓ from first infall into the MW-mass halo to present-day, that is, $(\ell(z = 0) - \ell_{\rm infall,MW})/\ell_{\rm infall,MW}$. Figure 4.3 shows this evolution versus $t^{\rm lb}_{\rm infall,MW}$ (left) and $M_{\rm star}$ (right). The median ℓ did not significantly change, *on average*, for satellites that fell in $\lesssim 10$ Gyr ago. However, earlier-infalling satellites systematically *increased* their angular momenta, by up to a factor of 2 on average, so angular momentum is not conserved for long periods in a dynamic cosmological halo environment, in part because the halo potential evolves on the same timescale as the orbit. We also stress that the typical $1-\sigma$ width of the distribution is ≈ 40 percent, so this represents the typical amount of dynamical scattering that a satellite experienced. Thus, the ensemble satellite population that fell in $\lesssim 10$ Gyr ago did not change in ℓ much, but any *individual* satellite's angular momentum changed by up to ≈ 40 percent, on average.

Figure 4.3 (right) shows that the median fractional difference in ℓ is minimal for satellites with $M_{\rm star} \lesssim 10^8\,{\rm M}_\odot$, but again, the $1-\sigma$ width of the distribution is $30-50$ percent. Higher-mass satellites experienced a stronger reduction in ℓ, by roughly 70 per cent, likely from the stronger dynamical friction they experienced.

Figure 4.2 thus highlights the strong dependencies of total velocity, E, and ℓ with the lookback time of infall and present-day galactocentric distance, and a lack of dependence on mass, except at $M_{\rm star} \gtrsim 10^8\,{\rm M}_\odot$. Our results in the middle column especially highlight the large distribution of these properties when selecting satellites at a given distance, and thus, we caution in solely interpreting the median values. Figure 4.3 highlights how earlier infalling and higher-mass satellites experienced larger changes in their specific angular momenta over time, as we explore below.

4.4.2. Orbital histories. In all but the most ideal conditions, the orbits of satellites change over time. Mechanisms such as the time-dependent (growing) MW-mass host potential, triaxiality of the host halo, dynamical friction, and satellite-satellite interactions can perturb satellite orbits. Because higher-mass satellites experience stronger dynamical friction, and because the MW-mass galaxy and host halo grew over time, a common expectation is that satellite orbits shrink over time,

such that their most recent pericentre generally should be their smallest (for example Weinberg, 1986; Taylor & Babul, 2001; Amorisco, 2017).

We now explore these expectations and examine trends for the infall times and pericentres in the orbital histories of satellites at $z = 0$. We investigate ensemble trends to help characterize and compare satellite populations in galaxies such as the MW, M31, and those in the SAGA survey. One should interpret these results for ensembles only, and not necessarily for individual satellites and their orbits, which we will explore further in future work.

Figure 4.4 caption: Lookback times to key events in the orbital histories of satellite galaxies at $z = 0$ versus their stellar mass, $M_{\rm star}$. Lines show the median, and the dark and light shaded regions show the 68th percentile and full distribution across the 13 hosts. We show a dotted horizontal black line at $z = 6$ to represent the end of reionization. **Top**: Lookback time of infall into the MW-mass halo, $t^{\rm lb}_{\rm infall,MW}$ (orange), and into any (more massive) halo, $t^{\rm lb}_{\rm infall,any}$ (black). For both infall metrics, higher-mass satellites fell in more recently, with $t^{\rm lb}_{\rm infall,any}$ ranging from ≈ 9 Gyr ago for lower-mass satellites to ≈ 2 Gyr ago for higher-mass satellites, and $t^{\rm lb}_{\rm infall,MW}$ ranging from ≈ 7.5 to 2 Gyr ago. Similarly, the 68th percentile and full distribution for both infall metrics span a similar time range, with the earliest infall times for any surviving satellite reaching nearly $t^{\rm lb}_{\rm infall,any} \approx 13.2$ Gyr and $t^{\rm lb}_{\rm infall,MW} \approx 12.5$ Gyr ago. Satellites with $M_{\rm star} \lesssim 3 \times 10^7$ M$_\odot$ fell into a more massive halo typically $1-2$ Gyr before falling into the MW-mass halo, thus experiencing 'group pre-processing'. **Bottom**: Lookback time to the most recent pericentre, $t^{\rm lb}_{\rm peri,rec}$ (green), and to the minimum pericentre, $t^{\rm lb}_{\rm peri,min}$ (purple), versus $M_{\rm star}$, for satellites with at least 1 pericentre. The median in $t^{\rm lb}_{\rm peri,min}$ decreases from $4-5$ Gyr ago for lower-mass satellites to $1-2$ Gyr ago for higher-mass satellites. Conversely, the median in $t^{\rm lb}_{\rm peri,rec}$ is roughly constant at $1-2$ Gyr ago across the entire mass range. At most masses, the minimum pericentre occurred $1-2$ Gyr *before* the most recent pericentre.

Figure 4.4 (top) shows the lookback times when satellites fell into the MW-mass halo, $t^{\rm lb}_{\rm infall,MW}$ (orange) or into *any* more massive halo, $t^{\rm lb}_{\rm infall,any}$ (black), as a function of satellite stellar mass, $M_{\rm star}$. By 'any more massive halo', we specifically mean any halo that is more massive than a

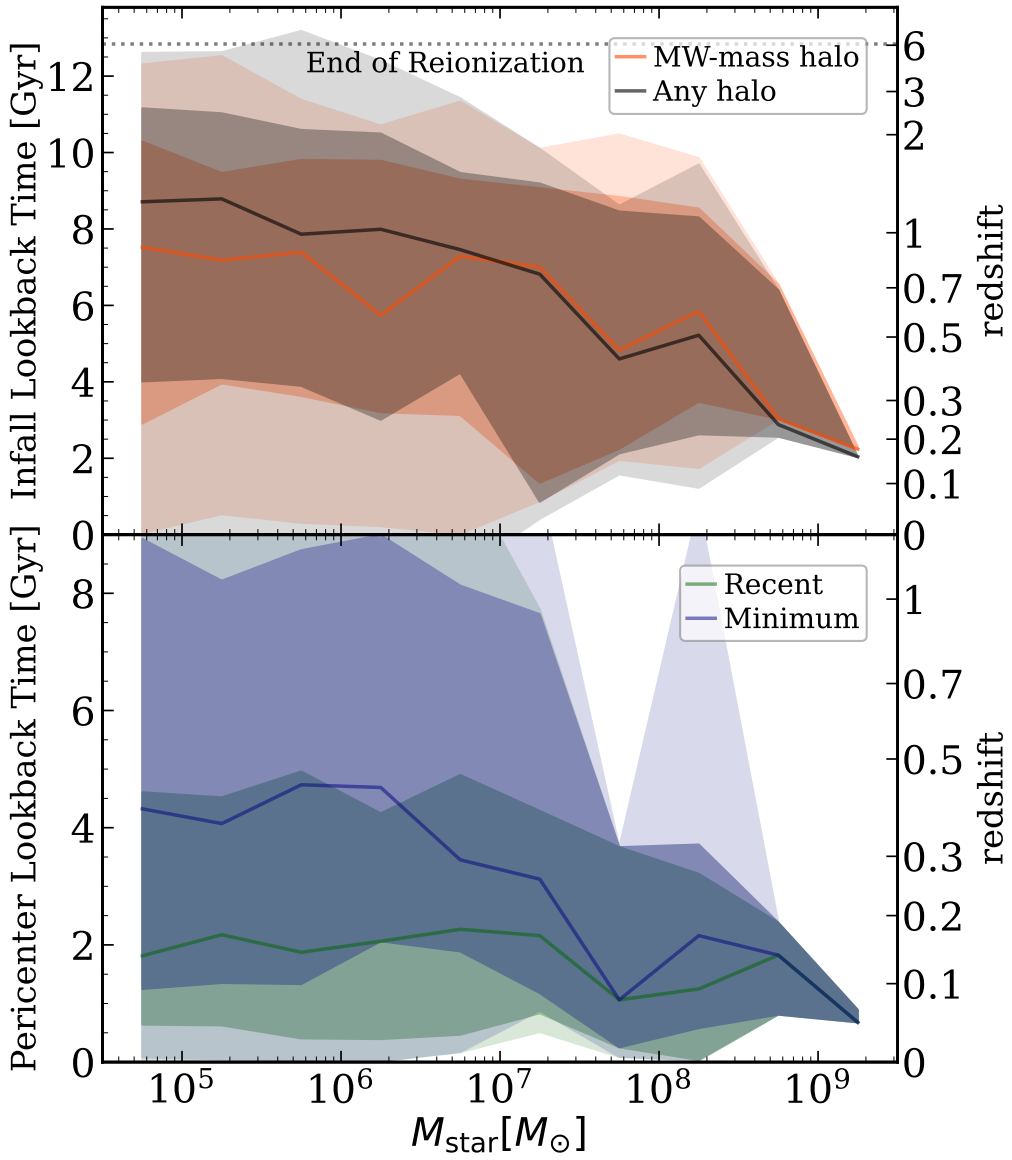

FIGURE 4.4. See figure caption in the text.

given satellite at the same time, which could either be a MW-mass host galaxy's halo or the halo of a massive central galaxy. We also show a horizontal dotted line to represent when the epoch of reionization ended (for example Robertson, 2022).

Galaxies form hierarchically, so early infalling galaxies, either into the MW-mass halo or any more massive halo, were less massive. Additionally, higher-mass satellites experienced stronger

dynamical friction, which caused their orbits to lose ℓ and merge with the MW-mass host more quickly. Infall lookback time into any more massive halo decreases from $t^{\rm lb}_{\rm infall,any} \approx 9\,{\rm Gyr}$ ago for satellites with $M_{\rm star} \approx 10^{4.75}\,{\rm M}_\odot$ to $t^{\rm lb}_{\rm infall,any} \approx 2\,{\rm Gyr}$ ago for $M_{\rm star} \approx 10^{9.25}\,{\rm M}_\odot$. At $M_{\rm star} \lesssim 10^7\,{\rm M}_\odot$, satellites typically fell into another (more massive) halo before they fell into the MW-mass halo. Above this mass, the limited range in mass between the satellite and the MW-mass halo does not leave room for an intermediate-mass host halo. Both the 68th percentile and full distribution for each infall metric span similar ranges, in particular, the 68th percentile ranges from $4 - 11\,{\rm Gyr}$ for satellites at $M_{\rm star} < 10^7\,{\rm M}_\odot$ and $1 - 9.5\,{\rm Gyr}$ at higher mass.

The earliest infalling satellites that survive to $z = 0$ fell into a more massive halo around $t^{\rm lb}_{\rm infall,any} \approx 13.2\,{\rm Gyr}$ ago ($z \approx 8.6$) and into the MW-mass halo at $t^{\rm lb}_{\rm infall,MW} \approx 12.5\,{\rm Gyr}$ ago ($z \approx 4.7$). Furthermore, $\approx 2/3$ of the satellites that fell into their MW-mass halo within the first 3 Gyr belong to the LG-like paired hosts, presumably because of the earlier assembly of halos in LG-like environments (see Santistevan et al., 2020). Similar to the analysis of dark matter-only simulations in Wetzel et al. (2015), no surviving satellites at $z = 0$ were within their MW-mass halo before the end of the epoch of reionization at $z \gtrsim 6$ (Robertson, 2022). In their analysis, less than 4 per cent of satellites were a member of a more massive halo as early as $\approx 13.2\,{\rm Gyr}$ ago ($z \approx 8.6$), near to when reionization was ≈ 50 per cent complete (see for example Faucher-Giguère, 2020), compared to < 1 per cent of the satellites in our sample. Thus, *reionization was finished before surviving satellites first became satellites of a more massive halo.* For a detailed study on satellite quenching after infall in the FIRE simulations, see Samuel et al. (2022b).

Roughly 37 per cent of satellites below $M_{\rm star} < 10^7\,{\rm M}_\odot$ were pre-processed before falling into their MW-mass hosts. Using the ELVIS DMO simulation suite of 48 MW-mass hosts, Wetzel et al. (2015) found that for pre-processed subhalos, the typical halo mass of the more massive halo they fell in was within $M_{\rm hosthalo} = 10^{10-12}\,{\rm M}_\odot$, with a median mass of $10^{11}\,{\rm M}_\odot$. From Figure 4.1, this corresponds to a median stellar mass of $10^{7-9}\,{\rm M}_\odot$. Wetzel et al. (2015) also determined that ~ 30 groups contributed all pre-processed satellites, for a typical MW-mass system, with each group

contributing between $2-5$ satellites. In our sample, satellites that were pre-processed typically fell into halos with $M_{\rm halo} = 10^{9.3-11}\,{\rm M}_\odot$ before falling into the MW-mass host halo, with a median mass at infall of $M_{\rm halo} = 10^{10.2}\,{\rm M}_\odot$. The stellar masses of galaxies hosted within these more massive halos ranged from $M_{\rm star} \sim 10^{6.4-8.9}\,{\rm M}_\odot$, with a median stellar mass of $M_{\rm star} = 10^{7.9}\,{\rm M}_\odot$. Thus, our results are consistent with Wetzel et al. (2015), however, we see a slightly less massive central halo mass.

Figure 4.4 (bottom) shows the lookback times to pericentres about the MW-mass host, for satellites that have experienced at least one pericentre. Some satellites have orbited their MW-mass host multiple times, so we show the lookback time to two pericentres: the most recent pericentre, $t^{\rm lb}_{\rm peri,rec}$ (green), and the minimum-distance pericentre, $t^{\rm lb}_{\rm peri,min}$ (purple). The mass dependence of $t^{\rm lb}_{\rm peri,rec}$ is weak. However, lower-mass satellites experienced earlier $t^{\rm lb}_{\rm peri,min}$ than higher-mass satellites, again because lower-mass satellites fell in earlier, when the MW-mass halo was smaller, so their overall orbit and pericentre distance was smaller than for higher-mass satellites that fell in at later times on larger orbits. Given that dynamical friction tends to shrink the orbits of satellites over time, particularly those with $M_{\rm star} \gtrsim 10^8\,{\rm M}_\odot$ or $M_{\rm halo,peak} \gtrsim 3\times10^{10}\,{\rm M}_\odot$, one might assume that, for satellites that have experienced multiple pericentres, the minimum pericentre should be equal to the most recent. However, for satellites with $M_{\rm star} \lesssim 3\times10^7\,{\rm M}_\odot$, the minimum pericentre occurred $1-3\,$Gyr earlier than the most recent pericentre, and the 68th percentile in $t^{\rm lb}_{\rm peri,min}$ spans a much larger range in lookback time, $0.25-9\,$Gyr, than $t^{\rm lb}_{\rm peri,rec}$, $0-5\,$Gyr. At $M_{\rm star} \gtrsim 3\times10^7\,{\rm M}_\odot$, where the typical number of pericentres experienced is only about 1 (see below), and where dynamical friction is more efficient, the two are more comparable. Thus, the naive expectation that low-mass satellite orbits remain relatively unchanged because they do not experience strong dynamical friction cannot explain the trends in Figure 4.4. Furthermore, although the medians between $t^{\rm lb}_{\rm peri,min}$ and $t^{\rm lb}_{\rm peri,rec}$ are comparable, the 68th percentile range (and full distribution) shows that differences between these pericentre metrics exist for satellites with $M_{\rm star} \gtrsim 3\times10^7\,{\rm M}_\odot$ also, implying that even the orbits of massive satellites can increase over time. The differences between the most recent and

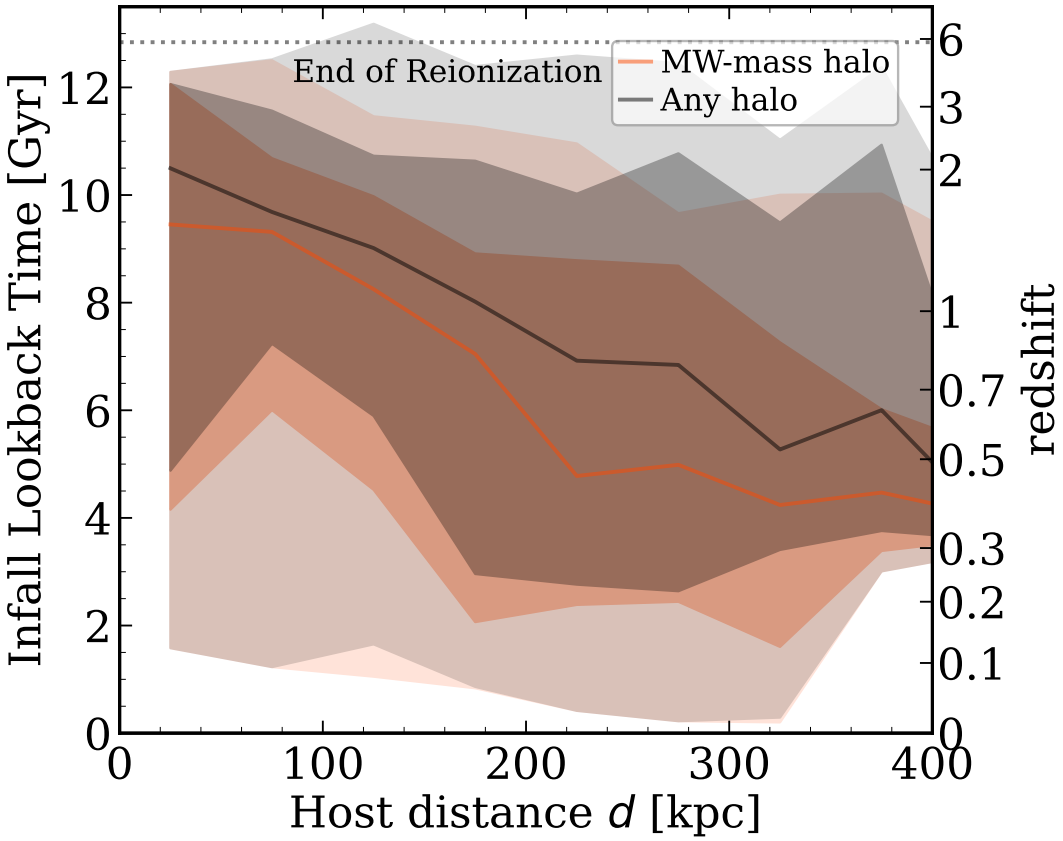

FIGURE 4.5. Similar to Figure 4.4 (top), but showing lookback time of infall into the MW-mass halo (orange) and into any (more massive) halo (black) versus present-day galactocentric distance, d. Lines show the median, and the dark and light shaded regions show the 68th percentile and full distribution across the 13 hosts. Satellites currently closer to the host typically fell in earlier, though with significant scatter. The lookback times of infall for the closest satellites are $10 - 10.75$ Gyr ago, and these times decrease to $4.5 - 5.5$ Gyr ago for the farthest. At any given d today, satellites fell into any more massive halo $0.5 - 2$ Gyr before they fell into the MW-mass halo. While d correlates with infall time, we emphasize the large scatter at a given d, which limits the ability of using d to infer infall time for any individual satellite.

minimum pericentres implies that some mechanism increases the pericentre distances of (especially lower-mass) satellites over time, as we explore below.

Figure 4.5 shows trends in $t^{\rm lb}_{\rm infall,MW}$ (orange) and $t^{\rm lb}_{\rm infall,any}$ (black) with present-day distance from the host, d. Satellites currently at closer distances typically fell into another halo earlier than

more distant satellites. The closest low-mass satellites fell into any more massive halo roughly $t^{\rm lb}_{\rm infall,any} \approx 10.5\,{\rm Gyr}$ ago, and this median time since infall decreases to 4.8 Gyr ago for satellites at $d \gtrsim R_{\rm 200m}$. Comparing this to the time since infall into their MW-mass halo, the median $t^{\rm lb}_{\rm infall,MW}$ is roughly $0.5 - 2\,{\rm Gyr}$ later across all distances. The range of both the 68th percentile and full distribution generally span similar lookback times, with $t^{\rm lb}_{\rm infall,any}$ offset to earlier times. The trend of more recent time since infall at larger d, is because at earlier times, the MW-mass halos were smaller and satellites fell in on smaller orbits.

Again, one should not interpret our results for individual satellites but rather for populations of satellites. Focusing on the full distribution, which extends across the full range in d, galaxies that fell into their MW-mass hosts between $t^{\rm lb}_{\rm infall,MW} \approx 3 - 9\,{\rm Gyr}$ ago currently orbit at all distances between $25 - 400\,{\rm kpc}$. Thus, although the median shows a clear trend with d, the range in the 68th percentile is $\gtrsim 2\,{\rm Gyr}$, which limits the ability to use the present-day distance of a given satellite to infer its time since infall.

Because of the mass dependence in satellite infall times in Figure 4.4, we checked for possible mass dependence of infall time with d by splitting the sample into lower-mass ($M_{\rm star} < 10^7\,{\rm M}_\odot$) and higher-mass satellites. The difference between $t^{\rm lb}_{\rm infall,any}$ and $t^{\rm lb}_{\rm infall,MW}$ exists at all satellite distances in the low-mass sample, and because the stellar mass function is steep, we saw nearly identical results to Figure 4.5. However, the higher-mass sample showed little to no difference between the two metrics, with times since infall ranging from $3.5 - 7\,{\rm Gyr}$ ago, because there were not many other more massive halos for higher-mass satellites to fall into before the MW halo.

Figure 4.6 caption: For satellite galaxies at $z = 0$, the number of pericentric passages about their MW-mass host, $N_{\rm peri}$ (top), and pericentre distance (bottom) versus their lookback time of infall into the MW-mass halo, $t^{\rm lb}_{\rm infall,MW}$ (left), present-day galactocentric distance, d (middle), and their stellar mass, $M_{\rm star}$ (right). We present the median (mean) trends in pericentre distance (number) in the curves, and show the 68th percentile and full distribution via the dark and light shaded regions, respectively. We include all satellites in the top row, but in the bottom row we include only

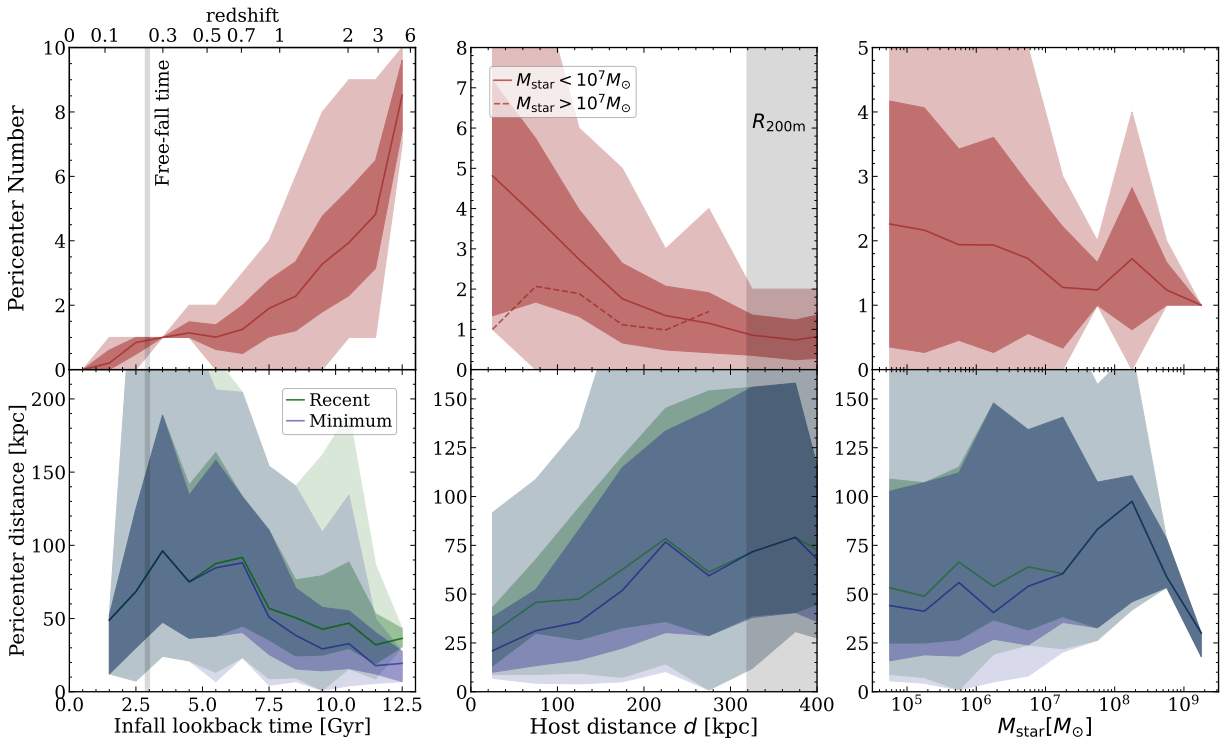

FIGURE 4.6. See the figure caption in the text.

satellites that have completed at least one pericentre. Similar to Figure 4.2, the vertical gray shaded bands show the free-fall time, t_{ff}, at R_{200m} (left) and the MW-mass halo radii, R_{200m} (middle). **Top row**: The mean N_{peri} increases with $t^{lb}_{infall,MW}$, where the most recently infalling satellites have not had enough time to complete a full orbit (left). This increase of N_{peri} for earlier infall is because satellites had more time to orbit, and because those that fell in earlier orbit at smaller distances. Lower-mass satellites experienced more pericentres (right), especially those currently at smaller distances (middle), given their earlier infall and the lack of strong dynamical friction acting on them. **Bottom row**: Satellites that fell into their MW-mass host earlier experienced smaller minimum pericentres, $d_{peri,min}$ (purple), and most recent pericentres, $d_{peri,rec}$ (green; left). Because lower-mass satellites and satellites that are more centrally located fell in earlier, they orbit with smaller $d_{peri,min}$ and $d_{peri,rec}$ (middle and right). However, higher-mass satellites also feel stronger dynamical friction that causes them to merge into the host on shorter timescales. The results

highlight how satellite orbits changed between the most recent and minimum pericentres, which often do not occur at the same distance, especially for satellites with $M_{\rm star} \lesssim 10^7\,{\rm M}_\odot$.

Figure 4.6 shows trends in the number of pericentric passages, $N_{\rm peri}$, about the MW-mass host (top row) and various pericentric distance metrics (bottom row), versus the time since infall into MW-mass halo, $t^{\rm lb}_{\rm infall,MW}$ (left), present-day distance from the MW-mass host, d (middle), and satellite stellar mass, $M_{\rm star}$ (right). Again, we include vertical gray shaded regions that represent the free-fall time at $R_{\rm 200m}$, $t_{\rm ff}$ (left column), and MW-mass halo $R_{\rm 200m}$ (middle), as reference values. When presenting trends in $N_{\rm peri}$, we include all satellites, including those that have not yet experienced a pericentric passage, but for trends in pericentre distance we only include satellites that have experienced at least one pericentre. Satellites that have not yet reached first pericentre comprise ≈ 7 per cent of all satellites.

The top left panel shows the expected trend of more pericentres for earlier $t^{\rm lb}_{\rm infall,MW}$. The mean $N_{\rm peri}$ is 0 for recently infalling satellites, and it rises to $N_{\rm peri} \approx 1$ and is flat across $t^{\rm lb}_{\rm infall,MW} \approx 2.5 - 5.5\,{\rm Gyr}$ ago, because this time interval is comparable to an orbital timescale. $N_{\rm peri}$ then rises rapidly with $t^{\rm lb}_{\rm infall,MW}$, reaching nearly 9 for the earliest-infalling satellites. We compared these trends for lower-mass versus higher-mass satellites and find no significant differences.

Figure 4.6 (top middle) shows the dependence of $N_{\rm peri}$ on d. Because we find significant differences in $N_{\rm peri}$ with $M_{\rm star}$ (top right), we split the sample into $M_{\rm star} < 10^7\,{\rm M}_\odot$ (solid) and $M_{\rm star} > 10^7\,{\rm M}_\odot$ (dashed). We choose this mass selection given that the lower-mass satellites experience a mean $N_{\rm peri} \geq 2$, while the higher-mass satellites have a mean of $N_{\rm peri} \approx 1$. Lower-mass satellites generally experienced more pericentres than higher-mass satellites at a given d, and the mean number decreases with d for lower-mass satellites from $N_{\rm peri} \approx 5$ to 1 for those near $R_{\rm 200m}$ (gray shaded region). Conversely, we do not find dependence on d for higher-mass satellites, with a mean value of $N_{\rm peri} \approx 1 - 2$ at all d, likely because of their more recent infall and the increased importance of dynamical friction on them.

Finally, $N_{\rm peri}$ declines weakly with $M_{\rm star}$, with a mean $N_{\rm peri} \approx 2.5$ at $M_{\rm star} = 10^{4.75}\,{\rm M}_\odot$ to $N_{\rm peri} \approx 1$ at $M_{\rm star} > 10^{9}\,{\rm M}_\odot$. Lower-mass satellites experienced more pericentres, because they fell in earlier (see Figure 4.4 top), and also because higher-mass satellites took longer to form and felt stronger dynamical friction that caused them to merge into their MW-mass host on shorter timescales. Over the full sample, the largest number of pericentres experienced is $N_{\rm peri} = 4$ at $M_{\rm star} > 10^{7}\,{\rm M}_\odot$, and $N_{\rm peri} = 10$ for lower-mass satellites.

The bottom row shows trends for both pericentre metrics: the pericentre with the minimum distance, $d_{\rm peri,min}$ (purple), and the most recent pericentre, $d_{\rm peri,rec}$ (green). In the idealized scenario we outline at the beginning of this subsection, an orbiting satellite's pericentre will remain unchanged or shrink over time because of dynamical friction. However, the panels above show that this often is not true; early infalling satellites can have larger subsequent pericentres.

The bottom left panel shows the median $d_{\rm peri,min}$ (purple) and $d_{\rm peri,rec}$ (green). Both $d_{\rm peri,min}$ and $d_{\rm peri,rec}$ are nearly identical for satellites that fell in $t^{\rm lb}_{\rm infall,MW} < 6\,{\rm Gyr}$ ago, with median values ranging from $60 - 100\,{\rm kpc}$. For earlier-infalling satellites, both distance trends slowly decrease from $\approx 100\,{\rm kpc}$ to $20 - 35\,{\rm kpc}$, where the most recent pericentre was roughly $5 - 20\,{\rm kpc}$ larger than the minimum. Thus, for sufficiently early-infalling satellites which spent longer amounts of time in the evolving MW-mass host halo, the orbits grew slightly over time, which is not expected given the assumption of unchanged orbits or shrinking orbits due to dynamical friction.

We also investigated how the first pericentre a satellite experienced depends on infall time and find qualitatively similar results to the other two pericentre metrics. Earlier-infalling satellites had smaller first pericentres than later-infalling satellites, and the first pericentres were smaller than the most recent ones. The first pericentres were also the minimum a satellite ever experienced in a majority (≈ 72 per cent) of satellites with $N_{\rm peri} \geq 2$. The only noticeable differences between the first pericentres and $d_{\rm peri,min}$ occurred for galaxies that fell in $\gtrsim 6\,{\rm Gyr}$ ago.

The bottom middle panel shows pericentre distance trends versus current d. As expected, both pericentre metrics increase with d, from $20 - 30\,{\rm kpc}$ for satellites that are currently closer, to

$70 - 80$ kpc for satellites near R_{200m} (gray shaded region). Satellites within $d \lesssim 225$ kpc often had recent pericentres that are larger than $d_{peri,min}$ by nearly $10 - 20$ kpc, so the orbits of these satellites grew. The pericentre metrics in both lower-mass and higher-mass satellites increase with d, but because lower-mass satellites fell in earlier than higher-mass satellites and completed more orbits, they largely drive the differences between $d_{peri,min}$ and $d_{peri,rec}$ (and $d_{peri,min}$ and $d_{peri,first}$) in the bottom left panel. We again highlight that the full distributions in both $d_{peri,min}$ and $d_{peri,rec}$ span a wide range at a given d, so even though the median trend increases with d, one should not directly apply our results to an individual satellite.

Finally, the bottom right panel shows that lower-mass satellites typically had smaller recent and minimum pericentres. However, at $M_{star} \gtrsim 10^{8.25}$ M$_\odot$, the median pericentre distances decrease, likely driven by the onset of efficient dynamical friction. Lower-mass satellites have smaller pericentre distances because they fell-in earlier when the MW-mass halo was smaller and less massive. The typical recent/minimum pericentre distance is $40 - 60$ kpc for satellites with $M_{star} \lesssim 10^7$ M$_\odot$, $60 - 100$ kpc for satellites with $M_{star} \approx 10^{7-8.25}$ M$_\odot$, and $\lesssim 100$ kpc for higher-mass satellites.

Because the mass of the host can determine the orbits of the satellites, we investigated potential differences between satellites in higher-mass and lower-mass host halos. At pericentre, satellites are deep in the potential near the galaxy, therefore, the stellar mass of the central galaxy could also correlate with our pericentre metrics. Specifically, we divided the sample in two by selecting the 6 MW-mass hosts with the higher M_{star} and 7 hosts with lower M_{star} (see Table 4.1) and examined their pericentre distances versus $t^{lb}_{infall,MW}$ and satellite M_{star}. We find no differences between the two samples versus M_{star}. Versus $t^{lb}_{infall,MW}$, the satellites in higher-mass host halos had slightly larger, although minimal, $d_{peri,min}$ and $d_{peri,rec}$.

The results in Figures 4.4-4.6 suggest a different evolution than expected for some satellites. Lower-mass satellites fell into their MW-mass hosts earlier, when the halo was smaller and less massive, so they complete more orbits than higher-mass satellites in this evolving potential and

orbit at smaller distances. Interestingly, the orbits of these lower-mass satellites can increase over time, presumably through the evolving global potential or interactions with other galaxies, which opposes the common expectation of shrinking orbits. However, given the 68th percentile ranges and the full distribution of the pericentre properties, differences between the most recent and minimum pericentres exist at *all* satellite masses, and not solely at low mass.

4.4.3. Satellites with growing pericentres. As we showed in Figures 4.4-4.6, the most recent pericentre that a satellite experienced is often not the minimum in terms of distance. We now investigate these cases in more detail and refer to satellites with $d_{\rm peri,min} < d_{\rm peri,rec}$ as having 'growing pericentres'.

Satellites with growing pericentres make up 31 per cent of all satellites (ranging from 23 – 46 per cent for a given host). Moreover, growing pericentres comprise the *majority* (67 per cent across all hosts, ranging from 50 – 86 per cent for a given host) of all satellites with $N_{\rm peri} \geq 2$. In other words, *for satellites with two or more pericentres, typically their most recent pericentre was not their closest encounter with their MW-mass host galaxy*. Figure 4.7 highlights this, showing the fraction of satellites with growing pericentres versus pericentre number. For satellites with $N_{\rm peri} \geq 2$, the growing pericentre population represents > 50 per cent of the total sample at any $N_{\rm peri}$, and in some cases, they represent the *entire* population at a given $N_{\rm peri}$. This fraction broadly increases with $N_{\rm peri}$, at least up to $N_{\rm peri} = 6$, where it represents all satellites, though the fraction fluctuates for $N_{\rm peri}$ above that.

Although we cannot directly check whether or not this is a temporary occurrence, we compared the most recent pericentre distance to the maximum pericentre a satellite experienced. For satellites that experienced more than 3 pericentres, we found that roughly 30 per cent of them experienced their maximum pericentre sometime between the minimum and most recent. Of these satellites, the fractional difference between their most recent pericentre distance, and their maximum, i.e. $(d_{\rm peri,recent} - d_{\rm peri,max})/d_{\rm peri,max}$, was between 1-54 per cent, with a median value of 17 per cent, and a 68th percentile range of 8-38 per cent. Thus, because this scenario happens in the minority

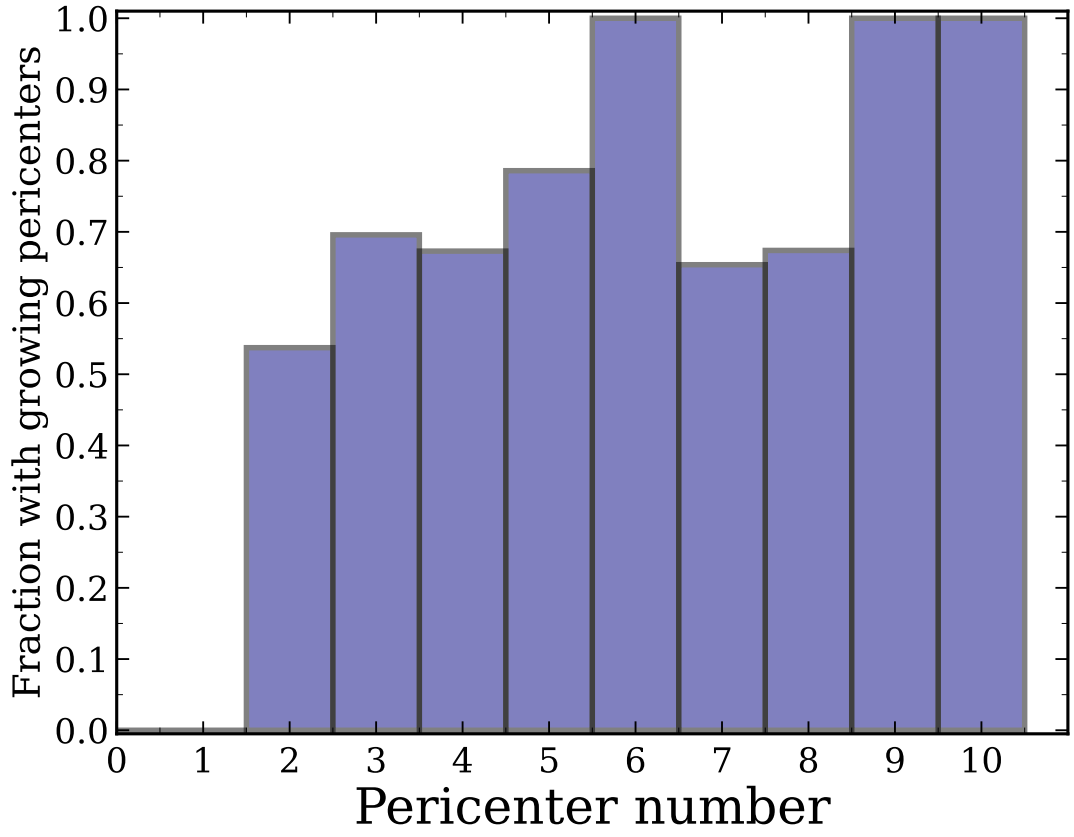

FIGURE 4.7. Fraction of satellite galaxies with growing pericentres, relative to all satellites that experienced the given number of pericentres, $N_{\rm peri}$. By definition, these satellites must have $N_{\rm peri} \geq 2$, which is why the fraction is zero for $N_{\rm peri} = 0$ and 1. *Satellites with growing pericentres represent the majority of all satellites with* $N_{\rm peri} \geq 2$.

of satellites, and because the median fractional difference is small, we argue that it is not merely a temporary occurrence. Furthermore, from the top right panel of Figure 4.6, satellites with more than 3 pericentres are generally lower-mass satellites, which will not strongly feel the effects of dynamical friction.

We confirmed that this population of satellites with growing pericentres is not sensitive to the choice for the center of the MW-mass host in computing satellite distances. Specifically, we examined these trends using the center of the host dark-matter halo (instead of the center of the stars in the host galaxy, as is our default). This results in only 7 additional satellites whose minimum

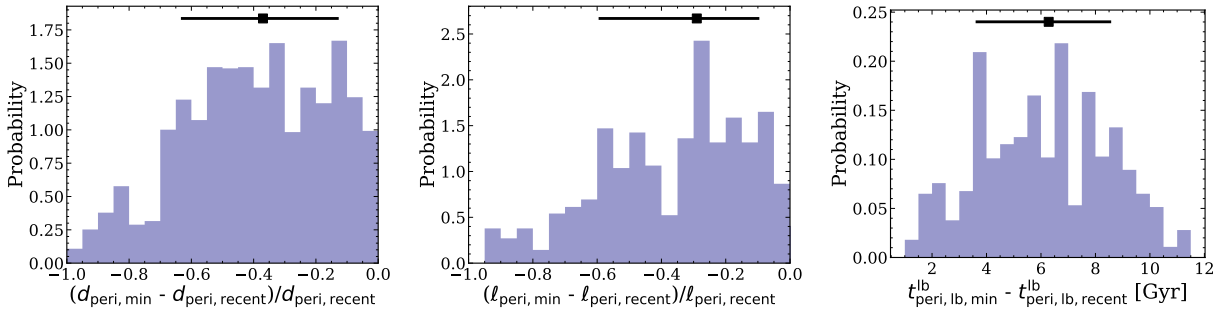

FIGURE 4.8. Probability distributions of changes in orbital properties for satellite galaxies with growing pericentres. This population comprises the majority (67 per cent) of all satellites that experienced multiple pericentres, or 31 per cent of satellites overall. Each panel shows the median value and 68th percentile via the black square point with error bars. **Left**: The fractional difference between the minimum and recent pericentre distances, $(d_{\rm peri,min} - d_{\rm peri,rec})/d_{\rm peri,rec}$. The median is −37 per cent. **Middle**: The fractional difference between the specific angular momenta at the minimum and recent pericentres, $(\ell_{\rm peri,min} - \ell_{\rm peri,recent})/\ell_{\rm peri,recent}$. The median is −29 per cent. **Right**: The difference between the lookback times to the minimum and recent pericentre, $t^{\rm lb}_{\rm peri,min} - t^{\rm lb}_{\rm peri,rec}$. The median is ≈ 6.3 Gyr, or $1-2$ orbits, given the free-fall time. *All metrics show significant differences between the most recent and the minimum pericentre for a given satellite.*

and most recent pericentres differ by more than 5 per cent, which represents only ≈ 4 per cent of all satellites with $N_{\rm peri} \geq 2$.

To quantify further the significance of this population, Figure 4.8 shows the probability distributions of the difference between key properties at the minimum and most recent pericentres for all satellites with growing pericentres. The left panel shows the fractional difference between the two pericentre distances, $(d_{\rm peri,min} - d_{\rm peri,rec})/d_{\rm peri,rec}$. As the black point shows, the median fractional difference is −37 per cent, with a 68th percentile range of −15 to −65 per cent.

Figure 4.8 (middle) shows the fractional difference in specific angular momentum at the minimum and most recent pericentres. Nearly all satellites with growing pericentres (> 95 per cent) experienced an increase in ℓ between the two pericentres; we do not show in Figure 4.8 the small percent with increased ℓ. The median fractional difference in ℓ is −29 per cent, with a range in the 68th percentile of −10 to −60 per cent.

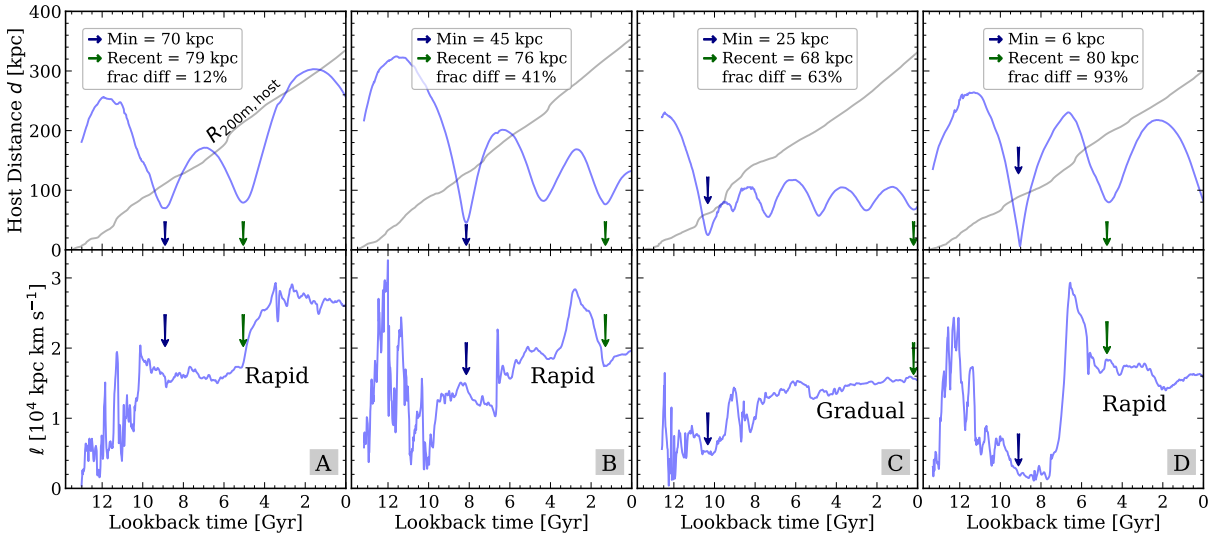

FIGURE 4.9. Orbital distance, d (top), and specific angular momentum, ℓ (bottom), for 4 representative satellites with growing pericentres, labeled A-D. For each satellite, we list its minimum and recent pericentre distances, along with the fractional difference between the two, and the purple and green arrows show when these events occurred. The top row also shows the growth of the MW-mass $R_{\rm 200m}$ (grey). The majority of satellites with growing pericentres (69 per cent) experienced $d_{\rm peri,min}$ within 1 Gyr after infall, in 72 per cent of satellites with growing pericentres the first pericentre was the minimum, and nearly 71 per cent of this satellite population orbited beyond the MW-mass host $R_{\rm 200m}$ after their first pericentre. The evolution of ℓ shows that the increase in pericentre distance can happen rapidly or gradually, but that it often does not happen near pericentre. *Approximately 30 per cent of cases show a sharp increase in ℓ during the first apocentre after infall, which suggests that these satellites may be interacting with other halos or non-axisymmetric features in the density field.*

Finally, Figure 4.8 (right) shows the difference between the lookback times of the minimum and most recent pericentres. These satellites have a wide range of time differences, with a median of ≈ 6.3 Gyr and 68th percentile range of $3.5 - 8.5$ Gyr. These are slightly longer than the typical orbital periods of these satellites, $2 - 5$ Gyr, as Figure 4.9 shows, the minimum and most recent pericentres do not always occur successively.

To provide more context, Figure 4.9 shows the orbits (host distance versus time) for four representative satellites with growing pericentres (top row), along with their specific angular momentum (bottom row), labeled A-D from left to right. We chose these four particular satellites

at random to span the entire possible range of fractional pericentre distances. The legends show the values of the minimum and most recent pericentres, along with the fractional differences between them; these four satellites range from 12 – 93 per cent. The arrows indicate when these pericentres occurred. For reference, we also show the MW-mass halo's $R_{\rm 200m}(t)$ (grey).

All four satellites experienced $d_{\rm peri,min}$ immediately after first infall, and the first pericentre was the minimum in 72 per cent of all satellites with growing pericentres. Furthermore, as Figure 4.9 suggests, 71 per cent of satellites with growing pericentres experienced a splashback phase of orbiting beyond $R_{\rm 200m,host}$ after their first pericentre. For comparison, among the population with $N_{\rm peri} > 1$ but $d_{\rm peri,min} = d_{\rm peri,rec}$, only 57 per cent experienced a splashback phase. So this suggest that orbiting beyond $R_{\rm 200m,host}$ is associated with a growing pericentre, at least in some cases.

As Figure 4.9 (bottom row) suggests, nearly all satellites whose pericentres grew also increased their specific angular momentum. By visually inspecting the histories of the full population, we find that this occurs in two broad ways: (1) steady, gradual increase in ℓ over time, which accounts for 45 per cent of all growing pericentres, and (2) rapid growth in ℓ near a pericentre or apocentre, which account for 53 per cent of the satellites. The remaining 2 per cent of satellites are the rare cases in which the pericentres increased from minimum to the most recent, but the angular momentum decreased. The fractional change in ℓ for these satellites is generally small, $\lesssim 6$ per cent. However, some of these satellites show clear signs of interactions with other galaxies, and fell in early ($\gtrsim 8.5\,$Gyr ago). For satellites in category (1), a time-dependent and/or triaxial host halo potential likely plays an important role, especially given that satellites with growing pericentres typically fell in early, when the shape of the host halo potential was changing more rapidly (for example Santistevan et al., 2020; Gurvich et al., 2022, Baptista et al, in prep). Satellite C in Figure 4.9 shows a relatively gradual increase in ℓ over time. We defer a more detailed investigation to future work.

For the growing pericentres in category (2), $4/5$ of the satellites experienced a rapid increase in ℓ near either a single apocentre, or some combination of them, with the first apocentre being the

most common. This is especially apparent in satellites B and D in Figure 4.9. The other satellites showed rapid increases in ℓ involving a pericentre that was not the minimum pericentre, much like in satellite A.

Because the fraction of splashback orbits is higher for satellites with growing pericentres compared to the remaining population with multiple pericentres, this suggests that perturbations at $d \gtrsim R_{\rm 200m,host}$ may play a key role in causing this population. This behavior is apparent in satellites B and D of Figure 4.9, where large spikes in ℓ occur near apocentres, some of which are beyond $R_{\rm 200m,host}$. These rapid increases in ℓ are caused by rapid increases in the tangential velocities, which typically were of order $\delta v \approx 30\,{\rm km\,s^{-1}}$.

Other satellite mergers with the MW-mass host also can significantly alter the global potential, resulting in orbit perturbations. We investigated correlations of both $t^{\rm lb}_{\rm peri,min}$ and $t^{\rm lb}_{\rm peri,rec}$ with the lookback times of mergers, with stellar mass ratios of $\gtrsim 1:100$, and did not find a clear correlation between these times. We also investigated correlations of these pericentre metrics with various metrics of host formation times including: the lookback times of when the host galaxy formed 90 per cent of its stellar mass (see Gandhi et al., 2022, for a table of values), the lookback times of when the host formed 10 per cent of its halo mass, and the lookback time of when the host galaxy's growth transitioned from being dominated by mergers to in-situ formation (Santistevan et al., 2020). We find no significant correlations with these formation metrics.

To investigate whether satellites with growing pericentres have biased orbits, both throughout their history and today, Figure 4.10 shows several orbital properties versus stellar mass, for satellites with growing pericentres, all satellites, and all satellites with $N_{\rm peri} > 1$ but with shrinking pericentres, that is, with $d_{\rm peri,min} = d_{\rm peri,rec}$. The top left panel shows the lookback time of infall into the MW-mass halo. Compared to the total sample, as expected, both sub-samples with $N_{\rm peri} > 1$ fell in typically $\gtrsim 1-2\,{\rm Gyr}$ earlier. However, among the population with $N_{\rm peri} > 1$, we find no significant differences between those with growing versus shrinking pericentres, so infall time does not correlate with having a growing pericentre.

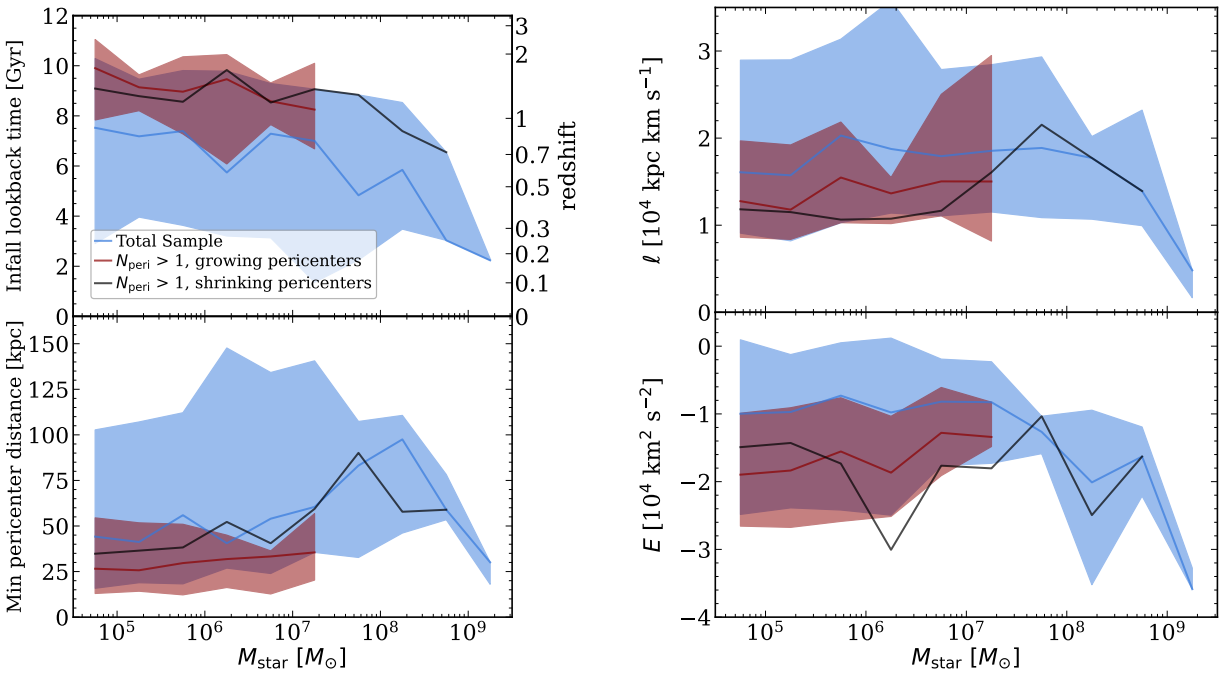

FIGURE 4.10. Orbital properties of satellite galaxies at $z = 0$, including the total sample (blue), those whose minimum and most recent pericentre distances are not the same ('$N_{\rm peri} > 1$, growing pericentres' in red), and those that experienced multiple pericentres but excluding the 'growing pericentre' population ('$N_{\rm peri} > 1$, shrinking pericentres' in black). Solid lines show median values, and shaded region shows the 68th percentile range for the growing pericentre and total population. **Top left**: Satellites that experienced multiple pericentres necessarily fell in earlier than the total sample, but among those with $N_{\rm peri} > 1$, the growing pericentre population is not systematically biased in infall time. **Bottom left**: Satellites with growing pericentres experienced $10 - 25$ kpc smaller $d_{\rm peri,min}$, while the remaining satellites with shrinking pericentres experienced $d_{\rm peri,min}$ more similar to the total sample. **Top right**: Because satellites with $N_{\rm peri} > 1$ fell into the MW-mass halo earlier than the total sample (when the host $R_{\rm 200m}$ was smaller), they have smaller ℓ today, and among these, satellites with growing pericentres have slightly higher ℓ today given their selection. More massive satellites with $N_{\rm peri} > 1$ have larger ℓ, closer to the total sample, because they fell in more recently. **Bottom right**: Because satellites with $N_{\rm peri} > 1$ fell into the MW-mass halo earlier than the total sample and orbit at smaller distances, they are more bound today. Satellites with growing pericentres have orbital energies today comparable to the other satellites with $N_{\rm peri} > 1$, though slightly higher at intermediate masses. In summary, satellites with growing pericentres fell in at comparable times to other satellites with $N_{\rm peri} > 1$, however, the satellites with growing pericentres orbit with smaller pericentre distances, and with slightly larger ℓ, which makes them unique.

Figure 4.10 (bottom left) shows the minimum pericentre distances, $d_{\rm peri,min}$. Although all three samples show similar behaviour to Figure 4.6, with increasing pericentre distance with increasing $M_{\rm star}$, the growing pericentre population is biased to the smallest $d_{\rm peri,min}$. The shrinking pericentre sub-sample is generally consistent with the total sample, with typical values spanning $d_{\rm peri,min} \approx 35 - 90\,{\rm kpc}$, while $d_{\rm peri,min}$ for growing pericentre satellites is $10 - 25\,{\rm kpc}$ smaller, ranging from $d_{\rm peri,min} \approx 25 - 35\,{\rm kpc}$. Thus, satellites with growing pericentres orbited closer to the host galaxy. Again, $\approx$ 30 per cent of satellites with growing pericentres experienced rapid increases in ℓ during their first apocentre or slightly after $d_{\rm peri,min}$. Likely, other important factors contribute to the larger differences between the pericentre metrics, such as the evolving MW-mass host potential, gravitational interactions with other satellite galaxies, and mergers, Figures 4.9 hints at, as we plan to explore in future work.

Figure 4.10 (top right) shows the specific angular momentum at $z = 0$. Because the growing and shrinking pericentre sub-samples fell into their MW-mass halo earlier, we expect them to have smaller ℓ. However, the growing pericentres having modestly higher ℓ at $z = 0$ at most masses, again reflecting that they have scattered to larger ℓ by today.

Figure 4.10 (bottom right) shows the specific orbital total energy, E. Consistent with their earlier infall, both sub-samples with $N_{\rm peri} > 1$ are on more bound orbits today than the total population, at least at $M_{\rm star} < 10^{7.25}\,{\rm M}_\odot$. Any systematic differences between the growing and shrinking pericentres are modest, given the scatter, so we conclude that there are no clear differences in specific orbital total energy at $z = 0$.

A satellite galaxy can undergo significant mass stripping when it orbits throughout the MW-mass host halo, especially when it is deepest in the host's potential at pericentre, and this drastic loss in the satellite's subhalo mass subsequently can affect its orbit. Thus, to better understand the origin of satellites with growing pericentres, including the timescales over which the orbits changed and potential dynamical perturbations near $d_{\rm peri,min}$, we compared the specific angular momentum and DM subhalo mass 200 Myr before and after the minimum pericentre. Near $d_{\rm peri,min}$, ℓ changed

by a much smaller amount (10 – 20 per cent) than the change in ℓ from the minimum to most recent pericentres ($\approx$ 40 per cent). The fractional mass lost near $d_{\rm peri,min}$ was also minimal ($\lesssim$ 7 per cent). Thus, in general, the orbital perturbations did not occur just near $d_{\rm peri,min}$, as also apparent in Figure 4.9.

Finally, we investigated the other orbital properties we presented in the previous figures (total velocity, pericentre lookback times, and the recent pericentre distances) for these sub-samples but find no compelling differences. We find no mass dependence to the fractional differences in pericentre distances or times in Figure 4.8, so satellites with growing pericentres exist similarly at a range of masses. Thus, even though higher-mass satellites experience stronger dynamical friction and have smaller orbital lifetimes, we find no mass dependence to whether or not a satellite has an orbit with growing pericentres. We find no strong correlation of the fractional distance or time metrics with either $d_{\rm peri,min}$, $d_{\rm peri,rec}$, or the lookback times that these occurred. Unsurprisingly, the fractional difference in pericentre distance increased slightly with earlier $t^{\rm lb}_{\rm infall,MW}$, given that satellites that orbited for a longer amount of time, had more time to experience changes in their orbits.

In summary, of all satellites that experienced $N_{\rm peri} \geq 2$, the majority (67 per cent) experienced a growing pericentre. The most recent pericentre distance is typically $\approx$ 37 per cent higher than the minimum experienced, which occurred $\sim$ 6 Gyr earlier. Interestingly, about half (45 per cent) of growing pericentres experienced a gradual increase in ℓ, presumably from a time-dependent and/or triaxial MW-mass host potential, and about half (53 per cent) experienced rapid growth in ℓ following either their first or minimum pericentres, during their first apocentres, or during multiple pericentre or apocentre events, which suggests a perturbation by another galaxy. Satellites with growing pericentres are more likely to have been splashback satellites, further suggesting perturbations at large distances. Furthermore, because we measure the orbits of these satellites relative to the MW-mass host galaxy, another effect may be perturbations to the center of mass of the host galaxy from mergers or massive satellites. Given these complexities and likely multiple

causes for the origin of satellite with growing pericentres, we defer a more detailed investigation to future work.

4.5. Summary & Discussion

4.5.1. Summary of Results. We investigated the orbital dynamics and histories of 473 satellite galaxies with $M_{\rm star} > 3{\times}10^4\,{\rm M}_\odot$ around 13 MW-mass galaxies from the FIRE-2 suite of cosmological simulations. Surprisingly, and in contrast to many (semi)analytical models of satellite evolution, most satellites that experienced multiple orbits experienced an increase in orbital pericentre and specific angular momentum, likely from interactions with the MW-mass host or other satellites. This highlights that satellite orbits do not always shrink and that angular momentum is not always conserved throughout a satellite's orbital history.

In summary, the topics that we presented in the Introduction and our corresponding results are:

1 *The relation of orbital properties of satellite galaxies at $z = 0$ to their orbital histories, including lookback times of infall, distances from the MW-mass host, and stellar masses.*

- Satellites that fell in earlier have lower orbital energies and specific angular momenta, *though with significant scatter*, because satellites that fell in earlier necessarily had to be on smaller orbits to be captured by the MW halo, and the MW-mass host potential continued to grow over time (Figure 4.2).
- Satellites closer to the host generally orbit with higher velocities, smaller specific angular momenta, and have more bound orbits, *though with significant scatter* (Figure 4.2). Total velocity, specific angular momentum, and specific orbital energy do not correlate with $M_{\rm star}$ except at $M_{\rm star} \gtrsim 10^8\,{\rm M}_\odot$, where dynamical friction is more efficient (Figure 4.2).
- *Specific angular momentum, ℓ, often is not conserved, even approximately, throughout a satellite's orbital history* (Figure 4.3). In particular, earlier-infalling satellites *increased* in ℓ since infall. More expectedly, higher-mass satellites decrease in ℓ, likely because of dynamical friction. The range of fractional changes in ℓ at smaller

$M_{\rm star}$ and later infall extends $\gtrsim$ 50 per cent. That said, the average ℓ across the full satellite population remains statistically unchanged since infall.

- Many lower-mass satellites were pre-processed before becoming a satellite of the MW-mass host. At $M_{\rm star} < 10^7\,{\rm M}_\odot$, 37 per cent fell into another more massive halo *before* falling into the MW-mass halo, typically ≈ 2.7 Gyr before (Figures 4.4, 4.5, A.2).
- *No surviving satellites were within the MW-mass halo during the epoch of reionization* ($z \gtrsim 6$), and less than 4 per cent were satellites of any host halo during this time, similar to Wetzel et al. (2015) (Figures 4.4 and A.2). Surviving satellites at $z = 0$ fell into the MW-mass halo as early as 12.5 Gyr ago, and into any host halo as early as 13.2 Gyr ago.
- Satellites at a given distance today experienced a large range of infall times into the MW-mass halo. Thus, one cannot infer a precise infall time based solely on a satellite's present-day distance alone, and the use of total velocity, specific angular momentum, or specific orbital energy alone is similarly limited (Figures 4.2, 4.5).

2 *Testing a common expectation that the orbits of satellite galaxies shrink over time, that is, that a satellite's most recent pericentric distance is the minimum that it has experienced.*

- Most satellites at $z = 0$ with $M_{\rm star} \lesssim 10^7\,{\rm M}_\odot$ experienced more than one pericentre, while more massive satellites experience only one (Figure 4.6), because of their later infall and dynamical friction.
- *Contrary to the expectation that satellite orbits tend to shrink over time, most satellites that experienced 2 or more pericentres have grown in pericentre distance.* Of all satellites with $N_{\rm peri} \geq 2$, 67 percent experienced a growing pericentre. This represents 31 per cent of all satellites.

- Typically, the minimum percienter was 37 per cent smaller than the most recent one, because the fractional specific angular momentum increased by 30 per cent (Figure 4.8). This minimum pericentre typically occurred $\sim 6\,\mathrm{Gyr}$ before the most recent one (Figure 4.8).
- Satellites with growing pericentres orbited closer to the host ($d_{\mathrm{peri,min}} = 24 - 35\,\mathrm{kpc}$) than those with shrinking pericentres.
- Perturbations at large distances likely contribute to these changes in satellite orbits, given the high fraction (71 per cent) of growing pericentres that were once a splashback satellite. However, we find no single dynamical origin: 53 per cent of satellites with growing pericentres experienced a large increase in ℓ during one or more apocentre, while 45 per cent experienced a gradual, steady increase in ℓ. This suggests that as the MW-mass host halo grows over time, this may help slowly torque the satellites to larger orbits, such that their subsequent pericentres increase. We leave a more detailed investigation of this to future work.

4.5.2. Inferring Infall Times from Present-day Properties. We presented various trends of present-day properties, such as total velocity, specific angular momentum, ℓ, specific energy, E, and distance from the host galaxy, d, with the lookback time of satellite infall, $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$. The median trends in these present-day properties often correlate with $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$. However, we stress that distribution of infall times at fixed property span a large range, limiting the ability to use a property like present-day distance to infer the infall time of a single satellite.

For example, in Figure 4.2, while the median specific energy decreases with increasing $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$, for a satellite with a specific energy of $E = -1 \times 10^4\,\mathrm{km^2\,s^{-2}}$, the 68 per cent range in $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$ is $1.5 - 10.5\,\mathrm{Gyr}$ ago. Similarly, although the median specific angular momentum decreases with increasing $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$, a satellite with $\ell = 2 \times 10^4\,\mathrm{kpc\,km\,s^{-1}}$ fell in $1 - 9\,\mathrm{Gyr}$ ago. Figure 4.5 shows that, for a satellite at 100 kpc today, $t^{\mathrm{lb}}_{\mathrm{infall,MW}} \approx 5.5 - 10.5\,\mathrm{Gyr}$, and for a satellite near the host virial radius, $d \approx 300\,\mathrm{kpc}$, it experienced $t^{\mathrm{lb}}_{\mathrm{infall,MW}} \approx 2 - 8\,\mathrm{Gyr}$ ago.

Furthermore, across Figures 4.2 and 4.5, at a given satellite total velocity, ℓ, E, or d, the *full* distribution of infall times spans ≈ 13 Gyr, nearly the age of the Universe. Thus, while these figure show trends in the median for a population of satellite galaxies, we caution that using any of one of these present-day properties for a single satellite will not precisely determine its infall time into the MW-mass halo. In future work, we will explore how precisely one can infer infall time using full 6D phase-space information, including knowledge about the host potential.

4.5.3. Comparison to Previous Work. First, we re-emphasize that these FIRE-2 simulations broadly reflect the observed population of satellites in the LG. Wetzel et al. (2016) and Garrison-Kimmel et al. (2019b) showed that their satellite stellar mass functions and internal velocity dispersions (dark-matter densities) broadly agree with the MW and M31. Samuel et al. (2020) showed that their radial distance distributions broadly agree with the MW and M31, and with MW-mass galaxies from the SAGA survey. Furthermore, Samuel et al. (2021) showed that, although uncommon, spatially or kinematically coherent planes of satellites exist in these simulations, similar to what is observed in the MW and M31. These benchmarks are important for motivating our analysis of their satellite orbits and histories.

Our results agree with Wetzel et al. (2015), who examined similar trends of satellite infall against both $M_{\rm star}$ and d, using the ELVIS suite of cosmological zoom-in DMO simulations, with abundance matching to assign stellar mass to subhalos across $M_{\rm star} = 10^{3-9}\,{\rm M}_\odot$. The mass and spatial resolution in these simulations were $1.9 \times 10^5\,{\rm M}_\odot$ and 140 pc, respectively, $\approx 5\times$ and $\approx 3\times$ larger than our baryonic simulations. They found that satellites typically first fell into any more massive halo $\approx 6.5 - 10$ Gyr ago and into the MW-mass halo between $\approx 5 - 7.5$ Gyr ago. These times since infall are consistent with the top panel of Figure 4.4 (and the results in Figure A.2) for lower-mass satellites, but the lookback times of infall for satellites at $M_{\rm star} \gtrsim 10^7\,{\rm M}_\odot$ in our results are more recent than in Wetzel et al. (2015): $\lesssim 6$ Gyr ago for both infall metrics. As we show in Appendix A.2, the addition of baryonic physics, especially the additional gravitational potential of the MW-mass galaxy, causes stronger tidal mass stripping and disruption, and because the

higher-mass satellites additionally feel stronger dynamical friction, we only see a few higher-mass satellites that happen to survive in our baryonic simulations. As a function of distance from the MW-mass host, Wetzel et al. (2015) also found that satellites experience first infall (into any more massive halo) $\approx 4-11$ Gyr ago and infall into their MW-mass halo $\approx 3-9$ Gyr ago, consistent with our results in Figure 4.5.

Rocha et al. (2012) used the Via Lactea II DMO simulation and found a strong correlation between satellite orbital energy and infall time (see also Fillingham et al., 2019; D'Souza & Bell, 2022). The authors suggest that satellites that are deeper in the gravitational potential at $z = 0$ often fell in earlier than satellites farther out and have more negative orbital energies. We find qualitatively consistent values and dependencies of infall time with d as in Rocha et al. (2012): satellites presently closer to the MW-mass galaxy fell in earlier and are on more bound orbits.

Bakels et al. (2021) used one of the N-body Genesis simulations, which have mass and spatial resolution of $7.8 \times 10^6\,\mathrm{M}_\odot$ and 1.1 kpc, respectively, which is $\approx 200\times$ and $\approx 30\times$ larger than the resolution in our simulations, to study the infall histories of satellites. They analyzed the orbits of (sub)halos of 2309 hosts with $M_{200\mathrm{c}} \geq 0.67 \times 10^{12}\,\mathrm{M}_\odot$ and found that roughly 22 per cent of all subhalos are on first infall, much larger than the $\approx 7-8$ per cent in our sample. Furthermore, Bakels et al. (2021) found that roughly 60 per cent of the splashback population of halos have yet to reach their first apocentre, and the majority (≈ 86 per cent) have only reached pericentre once, indicating that this population of satellites are on long-period orbits. Given the wide range in MW-mass $R_{200\mathrm{m}}$, if we select satellites that are currently beyond $300-400$ kpc to represent the splashback population, we find comparable results: over 95 per cent of the satellites have only experienced one pericentre so far, and the remaining 5 per cent have only experienced two pericentres. For the subhalos that have experienced pericentre at least once and are currently inside of the host's virial radius, Bakels et al. (2021) found that about half of them have only experienced one pericentre, and ≈ 30 per cent have not yet reached apocentre. Our result in the bottom left panel of Figure 4.6 is generally consistent with this, with a median pericentre number of $1-2$ for all satellites in our

sample. However, of satellites that experienced at least one pericentre, we find that $\gtrsim 2/3$ of them completed more than one, which suggests that the satellites in our simulations completed more orbits. Finally, Bakels et al. (2021) noted that roughly 95 per cent of the surviving subhalos were accreted since $z = 1.37$ (9.1 Gyr ago), where we generally see earlier infall: 95 per cent of our satellites fell into their MW-mass host since $z = 2.2$ (10.7 Gyr ago). Thus, the satellites that survive to $z = 0$ in our simulations fell in earlier resulting in the larger fraction that have completed more orbits. The differences in the first infall fractions, and the accretion times of satellite galaxies, between our results and Bakels et al. (2021) are likely because of the differences in resolution between the FIRE-2 and Genesis simulations. Because the Genesis simulations have DM particle masses $\approx 200\times$ larger, they necessarily resolve only more massive satellites.

Fattahi et al. (2020) used the cosmological baryonic zoom-in simulations of MW-mass galaxies from the Auriga project to investigate the $t^{\rm lb}_{\rm infall,MW}$ for surviving and destroyed low-mass galaxies, and their effect on the growth of the stellar halo. They also found that surviving satellites fell into their MW-mass halos more recently than the destroyed satellites, similar to the results in Panithanpaisal et al. (2021) and Shipp et al. (2022), who used the same 13 FIRE-2 simulations in our analysis to investigate stellar stream progenitors, their orbits, and their detectability. D'Souza & Bell (2021) also found similar results in their DMO satellite analysis. The analysis by Fattahi et al. (2020) shows similar results to the top panel in our Figure 4.4, with more massive satellites falling in more recently. At $M_{\rm star} = 10^6\,{\rm M}_\odot$ and $M_{\rm star} = 10^9\,{\rm M}_\odot$, the authors report average infall lookback times of $t^{\rm lb}_{\rm infall,MW} = 7.8\,{\rm Gyr}$ and $t^{\rm lb}_{\rm infall,MW} = 3.8\,{\rm Gyr}$. Our results are broadly consistent, though shifted to more recent times, where satellites at $M_{\rm star} = 10^6\,{\rm M}_\odot$ and $M_{\rm star} = 10^9\,{\rm M}_\odot$ fell into their MW-mass halo with mean infall lookback times of $t^{\rm lb}_{\rm infall,MW} \approx 6.5\,{\rm Gyr}$ and $t^{\rm lb}_{\rm infall,MW} \approx 2.9\,{\rm Gyr}$. We only have 3 satellites at $M_{\rm star} \sim 10^9\,{\rm M}_\odot$, smaller than the sample in Fattahi et al. (2020). The differences between the infall times between satellites in FIRE-2 and Auriga may arise from differences in the stellar mass - halo mass relation at low-mass (Grand et al., 2017; Hopkins et al., 2018). Furthermore, the way in which we both average satellite properties over the hosts may

contribute to the differences in infall time, given that host galaxies with larger satellite populations will skew the results.

Wetzel et al. (2015) concluded that, in the ELVIS suite of DMO simulations, no present-day satellites were within the MW-mass halo's virial radius during the epoch of reionization at $z \gtrsim 6$. This implies that, for any satellites whose star formation quenched during that time, the MW environment was not the driving factor, so the effects of the MW halo environment and cosmic reionization are separable, in principle. We similarly conclude that no satellites at $z = 0$ were within their MW-mass halo virial radius during reionization. Although our resolution is still finite, the trend in Figure 4.4 is relatively flat with mass, with no indication that it significantly increases for lower-mass satellites. Also, as we show in Appendix A.1, we find similar infall trends in subhalos down to $M_{\rm halo,peak} = 10^8\,{\rm M}_\odot$, which would host ultra-faint galaxies (Figure A.2). Recently, Sand et al. (2022) proposed that the ultra-faint galaxy Tucana B, whose nearest neighbor is $\approx 500\,{\rm kpc}$ away, was likely quenched in an isolated environment from reionization. It has an old ($\approx 13.5\,{\rm Gyr}$), metal-poor ([Fe/H] ≈ -2.5), stellar population, and has no recent star formation. Thus, because of its distance to any other massive galaxy, old stellar population, and lack of star formation, Sand et al. (2022) argued that Tucana B is an excellent candidate for a galaxy quenched by reionization. However, our results, and those in Wetzel et al. (2015), imply that no present-day satellites were within a MW-mass halo during reionization. Thus, selecting isolated galaxies *today* does not necessarily make them cleaner probes of the effects of reionization. Rather, satellites around MW-mass galaxies today provide similarly good candidates to study these effects.

Using the ELVIS DMO simulations, Wetzel et al. (2015) showed that many satellite galaxies first were pre-processed, for $0.5 - 3.5\,{\rm Gyr}$, before falling into their MW-mass halo; ≈ 30 per cent of satellites with $M_{\rm star} = 10^{3-4}\,{\rm M}_\odot$ were members of another group during their infall into a MW-mass halo, and this fraction decreases to ≈ 10 per cent at $M_{\rm star} = 10^{8-9}\,{\rm M}_\odot$. Any time before their infall into the MW-mass halo, ≈ 60 per cent of low-mass satellites were members of another more massive group, falling to ≈ 30 for high-mass satellites. Over our full sample of satellites, nearly

35 per cent fell into another more massive halo before falling into the MW-mass host, consistent with Wetzel et al. (2015). The fraction of pre-processed satellites in our results is also comparable to the DMO-based results from Li & Helmi (2008) who reported that $\approx 1/3$ of subhalos were pre-processed, though they selected subhalos down to $M_{\rm halo} \gtrsim 3 \times 10^6\,{\rm M}_\odot$ to probe subhalos that may not host luminous galaxies. Bakels et al. (2021) report that nearly half of all subhalos with $M_{\rm sub,acc}/M_{\rm host,200m} \sim 10^{-3}$ were pre-processed, and this ratio decreases for increasing subhalo mass. When specifically analyzing subhalos *on first infall*, Bakels et al. (2021) showed that as many as 40 per cent of subhalos were pre-processed prior to falling into their MW-mass halo, and this fraction increases for more massive host halos. More recently, D'Souza & Bell (2021) used the ELVIS DMO simulations to study the times since infall of subhalos with $M_{\rm halo,peak} > 10^9\,{\rm M}_\odot$ and how they were influenced by a massive merger ($> 1:10$). The distribution of times since infall for their surviving subhalos range $0 - 12\,{\rm Gyr}$, and the satellite $t^{\rm lb}_{\rm infall,MW}$ are peaked toward more recent values compared to the splashback population, which were accreted earlier. The full range of times since infall in our Figure 4.4 (and Figures A.2 and A.4) are consistent with the distribution in D'Souza & Bell (2021). Although D'Souza & Bell (2021) did not specifically focus on the first infall of subhalos into other more massive satellites/subhalos, they investigated group infall of satellites and showed that the distribution of time since infall clusters with the timing of the massive merger (and is slightly clustered with lower-mass mergers, $> 1:15$), with many subhalos becoming satellites of the massive merger $< 2.5\,{\rm Gyr}$ before it first crossed the MW-mass host radius.

Bakels et al. (2021) showed that after first infall, subhalos generally lose orbital energy and reach apocentres that are $\approx 0.8\times$ their turn-around radius, $r_{\rm ta}$, and all subsequent apocentres are typically comparable in distance. On the extreme ends, some subhalos gained or lost orbital energy and thus, reached larger or smaller subsequent apocentres, respectively, analogous to our satellites with growing pericentres. Regarding the subhalos that deviate strongly in their first apocentres and $r_{\rm ta}$, Bakels et al. (2021) found that nearly $\approx 2/3$ of the satellites with first apocentres $\gtrsim 3r_{\rm ta}$, and ≈ 80 per cent of the satellites that only reached $\lesssim 1/4r_{\rm ta}$, were pre-processed. Roughly $1/3$ of the

satellites with growing pericentres in our sample were pre-processed before falling into the MW, but they may also orbit outside of the more massive halo before falling into the MW halo. Thus, it is unlikely that pre-processing is the only driving factor in the origin and orbital evolution of satellites with growing pericentres.

Both Panithanpaisal et al. (2021) and Shipp et al. (2022) used the same 13 MW-mass galaxies in the FIRE-2 simulations that we use here to investigate stellar stream properties. Stellar streams form via disrupted low-mass galaxies or star clusters, however, before they completely disrupt, because they stretch throughout the halo we learn something about their initial orbits. Shipp et al. (2022) find that systems with smaller pericentres are more likely to form streams, and that the distribution of pericentres in the simulated streams are slightly smaller than the dwarf galaxies in our work. Furthermore, the authors suggest that not only are there differences in the orbital properties of present-day satellites and stellar streams, the orbits of streams with fully or partially disrupted progenitors differ as well, highlighting the complex evolution of low-mass stellar systems.

Finally, D'Souza & Bell (2022) explored uncertainties associated with orbit modeling using the ELVIS DMO simulations. They suggested that using simple parametric models for the MW-mass host (and recently accreted LMC-like galaxy) result in errors that are comparable to the 30 per cent uncertainty in the halo mass of the MW. They also extensively studied the errors associated with modeling the potential of a recently accreted LMC-like galaxy, the initial conditions of the satellites, and the mass evolution of the MW-mass halo, and they show that each comes with errors comparable to or less than the uncertainties in using simple parametric potentials.

Consistent with works like D'Souza & Bell (2022), our results highlight complications and limitations with idealized orbit modeling in a static, non-cosmological MW halo potential; most importantly, our results refute any expectation that the orbits of satellite galaxies always, or even generally, shrink over time. In Santistevan et al., in prep., we will use our simulations to pursue orbit modeling of individual satellite histories to compare with idealized orbit modeling in a static host potential.

Acknowledgements

We greatly appreciate discussions with Nora Shipp and Pratik Gandhi throughout the development of this paper. We are also thankful for interesting discussion with Andrey Kravtsov on the stellar mass - halo mass relation. IBS received support from NASA, through FINESST grant 80NSSC21K1845. AW received support from: NSF via CAREER award AST-2045928 and grant AST-2107772; NASA ATP grant 80NSSC20K0513; and HST grants GO-14734, AR-15809, GO-15902, GO-16273 from STScI. JS was supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2102729. RES gratefully acknowledges support from NASA grant 19-ATP19-0068, from the Research Corporation through the Scialog Fellows program on Time Domain Astronomy, from NSF grant AST-2007232, and from HST-AR-15809 from the Space Telescope Science Institute (STScI), which is operated by AURA, Inc., under NASA contract NAS5-26555. We ran simulations using: XSEDE, supported by NSF grant ACI-1548562; Blue Waters, supported by the NSF; Frontera allocations AST21010 and AST20016, supported by the NSF and TACC; Pleiades, via the NASA HEC program through the NAS Division at Ames Research Center.

CHAPTER 5

Modeling orbital histories of satellites of Milky Way-mass galaxies: testing static host potential models against cosmological simulations

5.1. Abstract

Understanding the evolution of satellite galaxies of the Milky Way (MW) and M31 requires modeling their full orbital histories across cosmic time. Many works that model satellite orbits incorrectly assume that the host halo gravitational potential is fixed in time and spherically symmetric. We explore and test the accuracy of static axisymmetric orbit modeling against the FIRE-2 cosmological simulations of MW/M31-mass systems. We fit the gravitational potential of each host at $z = 0$ and integrate the orbits of satellites back in time, comparing to the true orbit histories from the simulations. When a typical surviving satellite fell in (3.4 – 9.7 Gyr ago), the host halo mass and radius were typically 33 – 86 percent and 26 – 73 per cent of their values today. Orbital energy and specific angular momentum are not generally conserved and reach a typical uncertainty of 25 per cent around 4.8 and 10.3 Gyr ago, respectively. Recent orbit history properties, such as the timing and distance of the recent pericenter, typically deviate by 18 – 42 per cent, whereas events occurring farther back in time, such as the minimum pericenter distance or infall time deviate by 80 – 100 per cent. One must caution against overgeneralizing conclusions based solely on median trends as the variations within a population span $\gtrsim$ 75 per cent or more. Properties with smaller uncertainties include the timing and velocity of the recent pericenter, compared to the energy at infall and infall time at fixed $R_{200\mathrm{m}}$. Thus, our results are lower-limits given that we have near perfect knowledge of the simulation potentials.

5.2. Introduction

The satellite galaxies in the Local Group (LG) are the most rigorously studied low-mass galaxies to date, and will remain so for the foreseeable future because of their proximity to us. The dynamics and evolution of these low-mass galaxies encode rich information both about their past and the host environment in which they orbit. To better understand the orbit histories of satellites around Milky Way (MW)-mass host galaxies beyond the LG, we must first study and understand the orbits of satellite galaxies nearby. Several relevant questions about their orbit histories include: How has the mass of the MW changed over time, and how has it impacted the orbits of satellites over time? How well conserved are properties such as total orbital energy or angular momentum? Given a near-perfect representation of the host potential, to what extent can we accurately recover orbit properties, such as the most recent pericenter distance, if we keep the potential fixed over time, as is commonly done in orbit modeling studies?

The 6D phase-space coordinates of a satellite galaxy provide crucial information to study their orbits. One important, yet challenging, component in phase-space to measure for a satellite galaxy is its proper motion. Many studies used the Hubble Space Telescope (HST) to derive proper motion estimates for galaxies in the LG to estimate their star formation histories, determine companionship with the Large Magellanic Cloud (LMC), and to study the planar structure of satellites around both the MW and Andromeda (M31) (e.g. van der Marel et al., 2012; Kallivayalil et al., 2013; Sohn et al., 2020; Pawlowski & Tony Sohn, 2021). Recent HST treasury programs (e.g. GO-14734, PI Kallivayalil; GO-15902, PI Weisz; GO-17174, Bennet), obtained proper motion measurements for the remaining satellites with little data, and we now have kinematic information for all known satellites in the LG. The Gaia mission has been revolutionary in providing a wealth of data, such as positions, magnitudes, and proper motions for over 1 billion sources, including globular clusters and the satellites of the MW (Gaia Collaboration et al., 2018). Numerous studies use Gaia's kinematic data to study both group infall of satellites and to determine if galaxies are satellites of the LMC (e.g. Kallivayalil et al., 2018; Fritz et al., 2018a; Patel et al., 2020). Because proper

motion measurements improve with multiple observations, many studies now use HST and Gaia data in conjunction to reduce uncertainties in satellite galaxy proper motions (e.g. del Pino et al., 2022; Bennet et al., 2022; Warfield et al., 2023). In the advent of new science with the James Webb Space Telescope (JWST), we will soon be able to obtain proper motion measurements of even more distant or faint galaxies beyond the LG (e.g. see Weisz et al., 2023).

The global gravitational potential of the host galaxy primarily governs the orbits of its satellites. Thus, of upmost importance in deriving the satellite galaxy orbit histories in the LG is understanding the mass distribution of the MW and M31, or equivalently, their gravitational potential. In McMillan (2011, 2017), the authors decomposed the Milky Way into components such as the bulge, stellar/gaseous disk, and the dark matter halo, and fit observational data in these regions to analytic models to determine their best fit parameters. Some studies, such as Bovy & Rix (2013); Bovy et al. (2016a), focused on fitting and deriving parameters for the disk, such as the scale height and length, while other studies applied Bayesian modeling or used cosmological simulations in conjunction with observational data to estimate the total mass of the MW or M31 (e.g. Eadie et al., 2017; Patel et al., 2018; Eadie & Jurić, 2019; Patel & Mandel, 2022). Using constraints from globular cluster kinematics, and implementing a Monte Carlo Markov Chain method, Vasiliev (2019b) found that the virial mass of the MW was $M_{200\mathrm{m}} = 1.2 \times 10^{12}\,\mathrm{M}_{\odot}$, which is in line with the studies mentioned above and with the currently accepted virial mass in the literature of $M_{200\mathrm{m}} = 1 - 2 \times 10^{12}\,\mathrm{M}_{\odot}$ (e.g. Bland-Hawthorn & Gerhard, 2016). However, Patel & Mandel (2022) suggest the mass of M31 to be slightly more massive, $M_{200\mathrm{m}} = 2.85 - 3.02 \times 10^{12}\,\mathrm{M}_{\odot}$, from HST and Gaia proper motion measurements of satellite galaxies.

Satellite galaxies are not the only coherent stellar structures orbiting around the MW. Extended stellar streams around the MW provide a unique opportunity to study its mass distribution and potential, given that one can observe many stars along the length of the stream (e.g. Bonaca et al., 2014; Newberg, 2016; Bovy et al., 2016b; Panithanpaisal et al., 2022). Furthermore, the possible formation mechanisms of stellar streams is the disruption of satellite galaxies, or globular clusters

that may have been deposited by satellites in previous mergers. Therefore, studying the orbits of streams or globular gives insight to the possible future orbits of satellites that will eventually merge into their host galaxy (e.g. Ibata et al., 1994; Majewski et al., 1996; Bullock & Johnston, 2005; Price-Whelan et al., 2016, 2019; Panithanpaisal et al., 2021; Bonaca et al., 2021; Ishchenko et al., 2023). Most of the mass contained in the stellar halos around MW-mass galaxies also likely came from the complete disruption of a few lower-mass galaxies with the main host (e.g. Deason et al., 2016; Naidu et al., 2020). A number of studies have used both the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al., 2016) and the Southern Stellar Stream Spectroscopic Survey (S^5; Li et al., 2019) to discover these small systems and measure their kinematics (e.g. Shipp et al., 2018, 2019; Li et al., 2021, 2022). Comparisons with cosmological simulations in the recent study by Shipp et al. (2022) suggest that there may be currently undetected stellar streams around the MW, which may only be found through the upcoming Vera Rubin Observatory.

Using the phase-space information of satellite galaxies, in tandem with a model of the Galactic potential, many studies investigate satellite infall and orbital histories. A galaxy first becomes a satellite galaxy when it first crosses the virial radius of the more massive host, which can quench the lower-mass galaxy's star formation (for example Gunn & Gott, 1972; van den Bosch et al., 2008; Rodriguez Wimberly et al., 2019; Samuel et al., 2022b,a). As satellites reach their closest approach to the host galaxy at pericenter, they orbit deep within the potential in the more dense host CGM, and as such, feel strong tidal forces and may experience ram pressure stripping (e.g. Samuel et al., 2022a). On the other extreme, satellites spend more of their time at apocenter, farthest away from the host galaxy. Many studies use different models for the MW potential and numerically integrate the orbits for some of the MW's satellite galaxies to derive these orbit properties, however, the results depend strongly on the mass of the MW (e.g. see Gaia Collaboration et al., 2018; Fritz et al., 2018a,b). Another study by Fillingham et al. (2019) jointly used Gaia data with the Phat ELVIS (pELVIS) dark matter only (DMO) simulations, which include the gravitational effects of a central galaxy's disk, to match simulated satellites to observed satellites in 6D phase-space. They

then used the distribution of infall times of the matched simulated satellites to infer the infall times for 37 satellites of the MW, which ranged from $\approx 1 - 11$ Gyr ago, similar to other simulation-focused studies (e.g. Bakels et al., 2021; Santistevan et al., 2023). Deriving these infall and orbit history properties however, is generally tough given that we do not know precisely how the mass distribution of the MW has changed over time.

Before a low-mass galaxy falls into the MW-mass halo, it can become 'pre-processed' as a satellite of another intermediate-mass galaxy (for example D'Onghia & Lake, 2008; Li & Helmi, 2008; Wetzel et al., 2015; Deason et al., 2015; Kallivayalil et al., 2018; Jahn et al., 2019; Patel et al., 2020). Recent work not only suggests that the LMC has only recently fallen into the MW's halo and is currently near its first pericenter (e.g. Kallivayalil et al., 2013), but that it has a satellite population of its own (e.g. Deason et al., 2015; Kallivayalil et al., 2018; Patel et al., 2020). The stellar mass of the LMC is roughly 1/20 the stellar mass of the MW (van der Marel et al., 2002), therefore many studies now suggest that modeling its dynamical influence on other satellites is important in deriving realistic orbit histories (e.g. Garavito-Camargo et al., 2019; Patel et al., 2020; Garavito-Camargo et al., 2021; Samuel et al., 2021; Vasiliev et al., 2021; D'Souza & Bell, 2021; Correa Magnus & Vasiliev, 2022; D'Souza & Bell, 2022; Pace et al., 2022; Vasiliev, 2023). The recent interaction between some streams with the LMC have allowed astronomers to estimate the mass of the LMC (e.g. Erkal et al., 2018), where the current total mass estimate of the LMC based on this method by Shipp et al. (2021) is believed to be $M = 1.8 \times 10^{11}\,\mathrm{M}_{\odot}$.

Current observational programs such as the Satellites Around Galactic Analogs (SAGA) survey (Geha et al., 2017; Mao et al., 2021) are observing satellite populations around MW-mass galaxies. However, cosmological simulations of MW-mass galaxies provide similar opportunities. A number of studies focused on the infall and orbit histories of populations of satellite galaxies have used dark matter only cosmological simulations to understand how the dark matter in subhalos responds to pericentric events (Robles & Bullock, 2021), how subhalo orbits respond to various MW environments (Peñarrubia et al., 2002; Peñarrubia & Benson, 2005; Ogiya et al., 2021), pre-processing and

group accretion (Rocha et al., 2012; Wetzel et al., 2015; Li et al., 2020b; Bakels et al., 2021), and the counts and orbital distribution of subhalos (Barry et al., 2023). These, and many other studies, however, do not account for the effects that baryonic processes play in shaping galaxy formation and evolution (e.g. Brooks & Zolotov, 2014; Bullock & Boylan-Kolchin, 2017; Sales et al., 2022).

Many studies mentioned above rely on numerical tools to backward integrate the orbits of satellites, stellar streams, or globular clusters, such as *galpy* (Bovy, 2015), AGAMA (Vasiliev, 2019a), and Gala (Price-Whelan, 2017). However, orbit modeling studies often make approximations in their models by keeping the global mass/potential fixed over time (e.g. Patel et al., 2017; Fritz et al., 2018a; Fillingham et al., 2019; Pace et al., 2022), while varying other processes such as a changing MW center of mass or the inclusion of dynamical friction (Weinberg, 1986; Gómez et al., 2015; Patel et al., 2018; Garavito-Camargo et al., 2019, 2021), and understanding the limitations from these methods remains challenging (Klypin et al., 1999). To investigate the effects of different mass models on stellar streams dynamics in simulations, Arora et al. (2022) compared 4 models with and without time dependence and found that although most models conserve stream orbit stability, the only model that conserves stability over long periods of time is the time-evolving model. In a recent study by D'Souza & Bell (2022), the authors used 2 MW-mass hosts from the dark matter only simulations from the ELVIS suite of simulations to test how well orbit modeling recovers the true orbits of dark matter subhalos. The authors compared results from hosts with and without massive LMC-like satellites and showed that the model better recovers the more recent pericenters and apocenters when compared to the second or third-most recent. They also tested models in which they did not account for any mass growth of the MW-mass host or LMC-like companion and found varying degrees of uncertainty associated with each simple model. Although this study is robust in its methods and scientific inquiry, these simulations lack the inclusion of baryonic physics important in galaxy formation.

In this paper, we study the infall and orbit histories of the satellite galaxies in the cosmological, zoom-in simulations of MW-mass galaxies in the FIRE-2 simulations. We then model the mass

profiles of the MW-mass hosts in the simulations to within a few per cent at $z = 0$, backward integrate the orbits of satellites in a fixed representation of the $z = 0$ potential, and compare these results against those from the live, dynamic MW environments in the simulations. Thus, our goal is to investigate the strengths and limitations of static potential orbit modeling to provide insight/caution for this commonly used orbit modeling technique. Key questions that we focus on in this paper include:

i) How much has the MW-mass evolved during the time that these satellites first started orbiting in this environment and how does this affect our interpretation of results in static potential orbit modeling?

ii) Given our understanding of the orbital properties of satellite galaxies in the simulations, how well does the idealized model recover these properties, and are there properties that are easier to recover than others?

iii) How far back in time can we reliably integrate the orbits of satellites?

5.3. Methods

5.3.1. FIRE-2 Simulations. We use the cosmological zoom-in baryonic simulations of MW-mass galaxies in both isolated and LG-like environments from the Feedback In Realistic Environments (FIRE) project[1] (Hopkins et al., 2018). We ran these simulations using the hydrodynamic plus N-body code Gizmo (Hopkins, 2015), with the mesh-free finite-mass (MFM) hydrodynamics method (Hopkins, 2015), and the FIRE-2 physics model that includes several radiative heating and cooling processes such as Compton scattering, Bremsstrahlung emission, photoionization and recombination, photoelectric, metal-line, molecular, fine-structure, dust-collisional, and cosmic-ray heating across temperatures $10 - 10^{10}$ K (Hopkins et al., 2018). The FIRE-2 physics model also includes the spatially uniform and redshift-dependent cosmic ultraviolet (UV) background from Faucher-Giguère et al. (2009), for which HI reionization occurs at $z_{\rm reion} \sim 10$. Stars form in gas that is self-gravitating, Jeans unstable, molecular (following Krumholz & Gnedin, 2011), and dense

[1]See the FIRE project web site: `http://fire.northwestern.edu`

TABLE 5.1. Properties at $z = 0$ of the 13 MW/M31-mass galaxies/halos in the FIRE-2 simulations that we analyze, ordered by decreasing stellar mass. Simulations with 'm12' names are isolated galaxies from the Latte suite, while the others are from the 'ELVIS on FIRE' suite of Local Group-like pairs. Columns: host name; $M_{\rm star,90}$ is the host's stellar mass within $R_{\rm star,90}$, the disk radius enclosing 90 per cent of the stellar mass within 20 kpc; $M_{\rm 200m}$ is the halo total mass; $R_{\rm 200m}$ is the halo radius; and $N_{\rm satellite}$ is the number of satellite galaxies at $z = 0$ with $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$ that ever orbited within $R_{\rm 200m}$, totalling 493.

Name	$M_{\rm star,90}$ [$10^{10}\,{\rm M}_\odot$]	$M_{\rm 200m}$ [$10^{12}\,{\rm M}_\odot$]	$R_{\rm 200m}$ [kpc]	$N_{\rm satellite}$	Ref
m12m	10.0	1.6	371	47	A
Romulus	8.0	2.1	406	57	B
m12b	7.3	1.4	358	32	C
m12f	6.9	1.7	380	44	D
Thelma	6.3	1.4	358	34	C
Romeo	5.9	1.3	341	36	C
m12i	5.5	1.2	336	27	E
m12c	5.1	1.4	351	41	C
m12w	4.8	1.1	319	39	F
Remus	4.0	1.2	339	36	B
Juliet	3.3	1.1	321	40	C
Louise	2.3	1.2	333	34	C
m12z	1.8	0.9	307	26	F
Average	5.5	1.4	348	38	-

Note: Simulation introduced in: A: Hopkins et al. (2018), B: Garrison-Kimmel et al. (2019a), C: Garrison-Kimmel et al. (2019b), D: Garrison-Kimmel et al. (2017b), E: Wetzel et al. (2016), F: Samuel et al. (2020).

($n_H > 1000\,{\rm cm}^{-3}$), and represent single stellar populations, assuming a Kroupa (2001) initial mass function. Stars then evolve along stellar population models from STARBURST99 v7.0 (Leitherer et al., 1999), inheriting masses and elemental abundances from their progenitor gas cells. Other stellar feedback processes we implement in the FIRE-2 simulations include core-collapse and Ia supernovae, stellar winds, and radiation pressure.

We generated the cosmological zoom-in initial conditions at $z \approx 99$, which is early enough to capture both linear and non-linear growth of structure, and within periodic cosmological boxes of length 70.4 − 172 Mpc, which are large enough to avoid unrealistic periodic gravity effects on individual MW-mass hosts, using the code MUSIC (Hahn & Abel, 2011). We saved 600 snapshots for each simulation with time spacing of ≈ 25 Myr down to $z = 0$, assuming a flat ΛCDM cosmology with the following cosmological parameters consistent with Planck Collaboration et al. (2018): $h = 0.68 - 0.71$, $\sigma_8 = 0.801 - 0.82$, $n_s = 0.961 - 0.97$, $\Omega_\Lambda = 0.69 - 0.734$, $\Omega_m = 0.266 - 0.31$, and $\Omega_b = 0.0449 - 0.048$.

We analyze a similar set of galaxies as Santistevan et al. (2023), only we omit 'm12r', due to its relatively low stellar mass compared to the MW and because we were not able to accurately fit the enclosed mass profile (see Appendix B.1), and we include 'm12z' which was first introduced in Samuel et al. (2020). Our sample of galaxies are from both the Latte suite of isolated MW/M31-mass galaxies introduced in Wetzel et al. (2016) and the 'ELVIS on FIRE' suite of LG-like MW+M31 pairs, introduced in Garrison-Kimmel et al. (2019b). In Table 5.1, we list several properties, such as their stellar masses ($M_{\rm star,90}$), dark matter (DM) halo masses ($M_{\rm 200m}$) and radii ($R_{\rm 200m}$), and the number of satellite galaxies at $z = 0$ with stellar masses above $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$ ($N_{\rm satellite}$).

In the Latte suite of isolated MW/M31-mass galaxies, we chose halos at $z = 0$ with $M_{\rm 200m} = 1 - 2 \times 10^{12}\,{\rm M}_\odot$ with no other halos of similar mass within $5 \times R_{\rm 200m}$. We chose m12w to have LMC-mass satellite analogs near $z \sim 0$, and m12z to have a smaller DM halo mass at $z = 0$ (Samuel et al., 2020). Star particles and gas cells are initialized with masses of $7100\,{\rm M}_\odot$, however, because of stellar mass loss from supernovae, stellar winds, and radiation pressure, the average mass of star particles at $z = 0$ is closer to $\approx 5000\,{\rm M}_\odot$. The mass of DM particles is $3.5 \times 10^4\,{\rm M}_\odot$ within the zoom-in region. The gravitational softening lengths for star and DM particles are fixed at 4 and 40 pc (Plummer equivalent), respectively, comoving at $z > 9$ and physical thereafter. The gas cells use adaptive force softening down to 1 pc.

The selection criteria for each pair of halos in the 'ELVIS on FIRE' suite of LG-like galaxies is based on their individual masses ($M_{200\mathrm{m}} = 1 - 3 \times 10^{12}\,\mathrm{M}_\odot$), combined masses ($M_{\mathrm{tot}} = 2 - 5 \times 10^{12}\,\mathrm{M}_\odot$), their relative separation ($600 - 1000\,\mathrm{kpc}$) and radial velocities ($v_{\mathrm{rad}} < 0\,\mathrm{km\,s^{-1}}$) at $z = 0$, and their isolated environment (no other massive halos within 2.8 Mpc of either host center). The mass resolution in the 'ELVIS on FIRE' suite is $\approx 2\times$ better than the Latte suite, with initial star particle and gas cell masses between $\sim 3500 - 4000\,\mathrm{M}_\odot$.

The 13 simulated galaxies in our sample both display broadly consistent formation histories of similar MW/M31-mass galaxies, and exhibit comparable observational properties to the MW or M31, such as: MW/M31-like morphologies (Ma et al., 2017; Garrison-Kimmel et al., 2018; El-Badry et al., 2018b; Sanderson et al., 2020) that follow stellar-to-halo mass relations (Hopkins et al., 2018), realistic stellar halos (Bonaca et al., 2017; Sanderson et al., 2018a), and dynamics of metal-poor stars from early galaxy mergers (Santistevan et al., 2021). Each galaxy also hosts a satellite galaxy population with properties comparable to the satellites within the local Universe, such as: stellar masses and internal velocity dispersions (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a), radial and 3-D spatial distributions (Samuel et al., 2020, 2021), star-formation histories and quiescent fractions (Garrison-Kimmel et al., 2019a; Samuel et al., 2022b), and ensemble dynamic, orbital, and infall properties (Santistevan et al., 2023).

5.3.2. Halo/Galaxy Catalogs and Merger Trees. To generate the halo/subhalo catalogs at each of the 600 snapshots, we use the ROCKSTAR 6-D halo finder (Behroozi et al., 2013a) using DM particles only, and we use CONSISTENT-TREES (Behroozi et al., 2013b) to generate merger trees. There is no low-resolution DM particle contamination in any of the halos/subhalos that we analyze given each of the simulation's sufficiently large zoom-in volumes.

We will briefly review how we implement star particle assignment in post-processing here, however, we refer the reader to Samuel et al. (2020) for a more detailed description. First, we select star particles within $d < 0.8 \times R_{\mathrm{halo}}$, out to a maximum distance of 30 kpc, with velocities within $v < 2 \times V_{\mathrm{circ,max}}$ of the halo's/subhalo's center-of-mass (COM) velocity. After this selection, we

keep the star particles within $d < 1.5 \times R_{\rm star,90}$ of the (then) current member stellar population's COM and halo/subhalo center position, where $R_{\rm star,90}$ is the radius that encloses 90 per cent of the stellar mass. Then, we kinematically select the star particles with velocities within $v < 2 \times \sigma_{\rm vel,star}$ of the COM velocity of the member star particles, where $\sigma_{\rm vel,star}$ is the velocity dispersion of the current member star particles. Finally, we iterate on both the spatial and kinematic criteria until the halo's/subhalo's stellar mass converges to within 1 per cent. This also guarantees that the COM of the galaxy and its halo/subhalo are consistent with one another.

We use two publicly available analysis packages: HaloAnalysis[2] (Wetzel & Garrison-Kimmel, 2020a) for assigning star particles to halos and for reading and analyzing halo catalogs/trees, and GizmoAnalysis[3] (Wetzel & Garrison-Kimmel, 2020b) for reading and analyzing particles from Gizmo snapshots.

5.3.3. Selection of Satellites. We select satellites in the same manner as Santistevan et al. (2023). To summarize, we include all satellites at $z = 0$ above $M_{\rm star} > 3 \times 10^4\,{\rm M}_\odot$ that ever orbited within their MW-mass host halo's virial radius, $R_{\rm 200m}$. This stellar mass limit corresponds to roughly ~ 6 star particles, which reasonably resolves the total stellar mass (Hopkins et al., 2018). We include 'splashback' satellite galaxies that currently orbit outside of the MW-mass host's $R_{\rm 200m}$ near their apocenter, but are still gravitationally bound to their host (for example Wetzel et al., 2014). As Table 5.1 shows, the number of surviving satellites at $z = 0$ per host, including the splashback population, ranges from 26-57, and our sample totals 493 satellites. To avoid biasing our results to the hosts with larger satellite populations, we oversample the satellite properties so that each host contributes a near equal fraction of satellites to the total, to within 5 per cent; see Santistevan et al. (2023) for a more detailed discussion about this.

5.3.4. Orbit Property Calculations. We numerically integrate the orbits of satellite galaxies in custom mass/potential profiles we fit for each MW-mass host. In short, we model the mass

[2] https://bitbucket.org/awetzel/halo_analysis
[3] https://bitbucket.org/awetzel/gizmo_analysis

profiles of the hosts at $z = 0$ with a generalized form of the spherical Navarro-Frenk-White (NFW, Navarro et al., 1996) halo density profile, and two double exponential disk profiles, one for the inner disk and bulge contribution and another for the outer disk region. The median fit across all 13 MW-mass hosts is within ~ 5 per cent of the mass distributions in the simulations from $r > 10$ kpc out to the virial radius. Thus, *we test orbit modeling under the best of circumstances, with near perfect knowledge of the potential at present-day*. See Appendix B.1 for more details on the model and our derived parameters for each MW-mass host.

Next, we select the cylindrical positions (R, ϕ, Z) and velocities (v_R, v_ϕ, v_Z) of the satellites at the $z = 0$ snapshot and use this 6D phase-space information to initialize their orbits. We then use the galactic dynamics python package `galpy`[4] (Bovy, 2015), which allows users to define custom potentials, to backward integrate the satellite galaxy orbits for 13.8 Gyr.

We calculate the following properties for the satellites in our sample:

i) *Pericenter distance, time, and velocity:*

We calculate pericenters in the same manner as Santistevan et al. (2023), and we recapitulate the details here. We track the main progenitor of the satellites back in time through all 600 snapshots using the merger trees, and first confirm that the satellite is within the virial radius, $R_{200\mathrm{m}}$, of the MW-mass host halo at a given snapshot. Then, we find local minima in its galactocentric distance within a ± 20 snapshot window, which corresponds to ≈ 1 Gyr in time. Given the ~ 25 Myr time spacing between snapshots, we then fit a cubic spline to the distance, time, and velocity arrays in this snapshot window, and save the spline interpolated minimum distances, and the corresponding times and velocities at these pericenters.

As mentioned in Santistevan et al. (2023), we checked our pericenter calculations using a snapshot window of $\pm 4, 8, 10$ snapshots, and saw that our fiducial window of ± 20 snapshots best eliminates 'false' pericenters, i.e. cases where the pipeline finds a pericenter

[4] `http://github.com/jobovy/galpy`

in either numerical noise or short-lived perturbations in the orbits. Additionally, whether we center the distances on the satellite or MW-mass host galaxies or DM (sub)halos does not affect our results.

ii) *Apocenter distance:*

We calculate apocenters similar to the way we calculate pericenters. We first confirm that the satellite has orbited within $R_{200\mathrm{m}}$ of the MW-mass host halo before a given snapshot, however, we do not require it to be within $R_{200\mathrm{m}}$ at the snapshot of interest so that we may catch apocenters in the 'splashback' phase. We then find local maxima in the satellite's galactocentric distance within a similar ±20 snapshot window. Finally, we similarly fit a cubic spline to the distance and save these values.

iii) *Infall time:*

To calculate infall time, we simply ensure that the satellite is within $R_{200\mathrm{m}}$ at a given snapshot, and save the corresponding time that this first happens. In our model orbits, we calculate infall time in two different ways depending on how we treat $R_{200\mathrm{m}}$. The first method involves using the evolving $R_{200\mathrm{m}}$ from the simulations, and finding when the model orbit first crosses this distance, similar to how we calculate infall time for the satellites in the simulations. However, in our second method, we keep the present-day $R_{200\mathrm{m}}$ as a fixed quantity over time, i.e. $R_{200\mathrm{m}}(t_{\mathrm{lb}} = 0)$, where t_{lb} is lookback time in Gyr. We find the instances in which the model orbit first crosses within this fixed distance. In the case that the model orbit was always within $R_{200\mathrm{m}}(t_{\mathrm{lb}} = 0)$, we set the infall time equal to the beginning of the simulation, 13.78 Gyr ago.

iv) *Eccentricity:*

We approximate orbital eccentricity using the following relation:

$$e \approx \frac{d_{\mathrm{apo}} - d_{\mathrm{peri}}}{d_{\mathrm{apo}} + d_{\mathrm{peri}}} \tag{5.1}$$

where $d_{\rm apo}$ and $d_{\rm peri}$ are the apocenter and pericenter distances, respectively. The time between pericenter to apocenter is approximately half an orbit, thus, we calculate eccentricity at each half orbit, i.e. using each adjacent pericenter-apocenter pairs.

v) *Orbital period:*

We approximate the orbital period by simply calculating the time between adjacent pericenters. We checked how this compares to the timing between adjacent apocenters and found consistent results.

vi) *Specific orbital energy:*

We take the specific orbital energy of a satellite to be the sum of the kinetic and potential energy. In Santistevan et al. (2023), we normalized the potential energy such that the total energy is defined to be 0 at present-day for a satellite on a circular orbit at the virial radius. However, since we only look at differences in total energy in this paper, the absolute normalization at present-day is not important. The simulations provide the gravitational potential of satellites at different distances and we ensure that the potential is normalized properly across time with

$$U_{\rm sat,norm}(r, t_{\rm lb}) = U_{\rm sat,raw}(r, t_{\rm lb}) - U_{\rm host,raw}(r = 500\,{\rm kpc}, t_{\rm lb}) - \frac{G \times M(< 500\,{\rm kpc}, t_{\rm lb} = 0)}{500\,{\rm kpc}} + \frac{G \times M(< 500\,{\rm kpc}, t_{\rm lb})}{500\,{\rm kpc}} \tag{5.2}$$

where $U_{\rm sat,raw}(r, t_{\rm lb})$ is the raw satellite potential from the simulations, $U_{\rm host,raw}(r = 500\,{\rm kpc}, t_{\rm lb})$ is the raw potential for particles within a spherical shell at $r = 500 \pm 5\,{\rm kpc}$ around the MW-mass host, and the last two terms are the analytic gravitational potentials, $G \times M(< r)/r$, for an mass enclosed within 500 kpc at present-day and any given lookback time, respectively. Because satellites typically orbit as far as the 'turn-around' radius, which we approximate as $\sim 1.5 \times R_{\rm 200m} \approx 500\,{\rm kpc}$ from a spherical collapse model, we

assume that the enclosed mass out at 500 kpc is not changing very much. Thus, because

$$a = \frac{G \times M(< r)}{r^2} = \frac{\partial \Phi}{\partial r} \tag{5.3}$$

where a is the acceleration, and $\partial\Phi/\partial r$ is the partial derivative of the potential, Φ, with respect to r, if we integrate the right-hand side of the equation to get Φ, we get

$$\Phi = \int_{500\,\mathrm{kpc}}^{\infty} \frac{G \times M(< r)}{r^2} dr. \tag{5.4}$$

Assuming the enclosed mass does not change significantly beyond 500 kpc, the integral results in $\Phi = G \times M(< 500\,\mathrm{kpc})/500\,\mathrm{kpc}$. The last three terms in Eq. 5.2 ensure that the potential is properly normalized across different snapshots.

Finally, in the plots where we look at differences in total energy, we divide the differences by $U_{200\mathrm{m},0} \equiv G \times M(r < R_{200\mathrm{m}}, t_{\mathrm{lb}} = 0)/R_{200\mathrm{m}}$. Because each host has different values for $M_{200\mathrm{m}}$ and $R_{200\mathrm{m}}$, this ensures that the satellites in each system are treated in a similar manner. We note however, that 'm12z', 'Romulus', and 'Remus' do not have stored potentials in the simulations, thus, we exclude them from any plots where we compare energies.

Finally, we checked how these properties varied when using satellite coordinates that are aligned with the MW-mass host disk, and those that are not. The default coordinates are defined with respect to the moment of inertia tensor of the MW-mass disk. To convert these coordinates, we rotate the cylindrical coordinates with the rotation matrix associated with the moment of inertia tensor of the disk calculated at $z = 0$. Because satellites at pericenter are closest in their approach to the MW-mass host disk, we investigated how these pericenters differed between the aligned vs unaligned coordinates. In general, the median pericenter difference between the two coordinates was effectively zero ($\sim 10^{-6}$ kpc), and the 68th percentile range of values was ~ 100 pc. The largest differences we saw in derived pericenters was ~ 3 kpc. Because pericenters are the orbit property most likely to be affected by the host disk, and because we only see small differences between the

two methods, our results are not strongly dependent on whether satellites are aligned with the disk or the DM halo. Thus, hereon our results use coordinates that are aligned with the disk.

5.4. Results

5.4.1. Growth of the Milky Way-mass Hosts. Many studies integrate satellite orbits using a model of the static, symmetric MW potential at $z = 0$. However, this is incorrect given what we know about hierarchical galaxy formation. Before we derive and compare results from static orbit modeling, it is important to understand the extent to which the mass of the MW-mass hosts change over time. Here, we investigate and quantify the mass evolution of the MW-mass hosts in our sample on both long and short timescales, and how this may depend on distance.

We define virial properties of a (sub)halo at $R_{200\mathrm{m}}$, i.e. at the radius that encloses 200× the mean matter density of the Universe. In Figure 5.1, we present the evolution of both the total (baryonic plus dark matter) virial mass, $M_{200\mathrm{m}}$ (left), of the hosts, and their virial radii, $R_{200\mathrm{m}}$ (right). In each panel we show the median trend in the solid black line, along with the 68th and 95th percentiles of the sample in the dark and light shaded regions, respectively. The left-hand axes in each panel show the virial properties normalized to the present-day ($t_{\mathrm{lb}} = 0$, $z = 0$) values, while the right-hand axes show the physical units corresponding to the median. Because low-mass galaxies become satellites of MW-mass galaxies by orbiting within the dark matter halos of their more massive hosts, in both panels, we show the median satellite infall lookback time into their MW-mass hosts, $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$, in the grey vertical line, and the 68th percentile range of values with the light grey shaded region.

In the left panel, the median MW-mass host qualitatively grows in mass faster at earlier times, and around ~ 4 Gyr ago, the fractional growth slows down. These galaxies likely increase in mass faster early in the simulation where mergers with low-mass galaxies are more common than at later times. As Santistevan et al. (2020) showed, the fractional stellar mass growth of these MW-mass hosts is consistent with other abundance matching and dark matter only studies (for example Hill et al., 2017; Behroozi et al., 2013b). The 68th percentile range of $t^{\mathrm{lb}}_{\mathrm{infall,MW}}$ is between $3.4 - 9.7$ Gyr

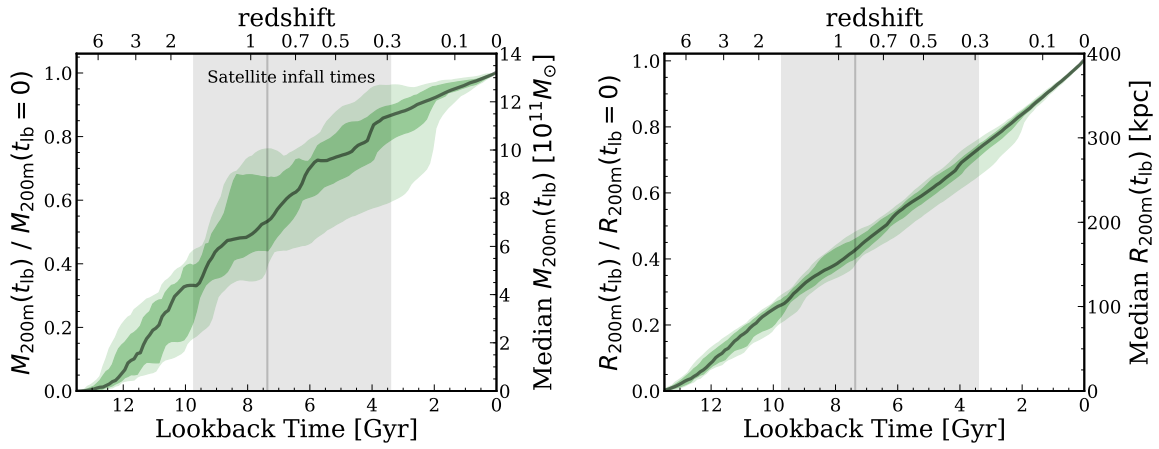

FIGURE 5.1. **Left**: Total (baryonic plus dark matter) mass of a MW/M31-mass halo, M_{200m}, enclosed within R_{200m} as a function of lookback time, t_{lb}. The left axis shows M_{200m} at t_{lb} relative to M_{200m} today ($t_{lb} = 0$ Gyr), while the right axis converts this to the median M_{200m} across our sample. The black line shows the median, and the dark and light shaded regions shows the 68th and 95th percentiles across our 13 hosts. The vertical grey line and shaded region show the median and 68th percentile range of the lookback times of infall for surviving satellites. The typical host halo had 54 per cent of its final M_{200m} when a typical satellite fell in ≈ 7.4 Gyr ago, so M_{200m} nearly has doubled over that time, with even more growth experienced by earlier-infalling satellites. **Right**: Same, for the growth of the MW/M31-mass halo virial radius, R_{200m}. R_{200m} shows nearly linear growth with time over the last ≈ 12 Gyr. The typical host halo had 43 per cent of its final R_{200m} when a typical satellite fell in ≈ 7.4 Gyr ago, again highlighting significant change over the orbital history of a typical satellite.

ago, and this range of lookback times qualitatively corresponds to the largest range of M_{200m} values across the hosts.

In the right-hand panel of Figure 5.1, we see that the growth of R_{200m} is nearly linear across all lookback times. The widths of the 68th and 95th percentiles are relatively small at fixed radius, highlighting the self-similarity in the growth of the size of these hosts. In comparison, at fixed mass, the width of the 95th percentile in M_{200m} spans as large as ~ 6 Gyr, whereas at fixed radius it spans 1.5 Gyr at most. This suggests that although the sizes of these galaxies evolve in a nearly similar manner, their mass growth is more stochastic given their unique formation histories.

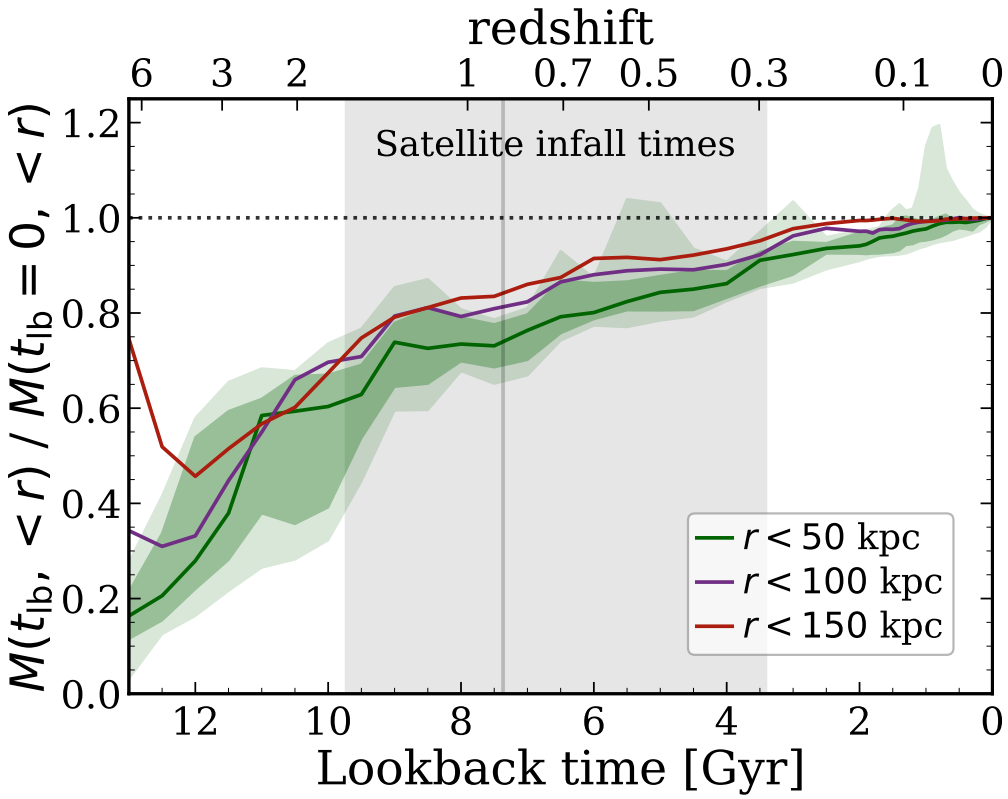

FIGURE 5.2. Total mass of the MW/M31-mass host within a fixed physical radius, r, normalized to the value today, as a function of lookback time, $t_{\rm lb}$. We show the median mass ratio within r = 50, 100, and 150 kpc, and the 68th and 95th percentiles for r < 50 kpc. This 'physical' mass growth is less significant than that of $M_{\rm 200m}$ (based on an evolving $R_{\rm 200m}$) in Figure 5.1. However, because of baryonic cooling, the fractional mass growth is larger at smaller physical radii. The mass growth reverses at $t_{\rm lb} \gtrsim 11$ Gyr, when a given radius experienced its initial collapse from the Hubble expansion. For context, the median pericenter distance experienced by surviving satellites is $\approx$ 50 kpc. A typical satellite fell in 3.4 – 9.7 Gyr ago, when the typical enclosed mass was $\approx$ 61 – 93 per cent of its value today.

During typical satellite infall times, the median MW-mass hosts were only $\sim$ 33 – 86 per cent of their z = 0 total masses, and 26 – 73 per cent of their z = 0 radii. Thus, since these galaxies first became satellites, the MW-mass hosts grew considerably in mass and size, which likely affected the orbits of these satellites.

In Figure 5.1, we investigated the enclosed mass within the entire evolving virial region, $R_{200\mathrm{m}}(t)$, but to better investigate how the inner regions of the host grew, in Figure 5.2, we discuss trends in the enclosed mass within fixed spherical apertures across time. In Figure 5.2, we present the ratio of the enclosed mass within r at a given lookback time, t_{lb}, relative to the enclosed mass at present-day, i.e. $M(t_{\mathrm{lb}}, < r)/M(t_{\mathrm{lb}} = 0, < r)$. We show the median ratios of enclosed mass within $r < 50$, 100, and 150 kpc in the green, purple, and red lines, respectively. In the dark and light green shaded regions, we show the 68th and 95th percentiles for the $r < 50$ kpc selection only, for visual clarity.

13 Gyr ago, the enclosed mass ratio within $r < 50$ kpc was only 15 per cent of what it is today. The enclosed mass ratio was $\sim 62 - 91$ per cent during the typical satellite infall times, which skew to larger values than the mass ratios within the virial radius in Figure 5.1. Thus, the inner region of the halo formed a majority of its total mass by the time many of these satellites fell in for the first time; only ~ 17 per cent of satellites fell in earlier than ~ 9.7 Gyr ago. Within the 95th percentile, we see a few temporary spikes in the enclosed mass, which likely result from nearby orbiting satellites.

The median enclosed mass within $r < 100$ kpc and $r < 150$ kpc skew to larger values at nearly all lookback times. Thus, at any given time, larger distances encompass more mass, and the median trends suggest that the total mass within larger fixed apertures do not change as much as the inner regions do over time. Put differently, these larger distance selections formed a given fraction of their present-day mass earlier than the smaller selections. The main galaxy represents a larger fraction of the enclosed mass at smaller distances, and as the galaxy grows over time, gas continues to cool and fall into these inner regions, which is why smaller distance selections form a fraction of their mass later. The distances we show in this plot are *physical* distances, not co-moving, so the enclosed mass within 100 and 150 kpc at $t_{\mathrm{lb}} > 11$ Gyr increases, unlike the 50 kpc selection. At such early times, some of the mass within 100 and 150 kpc is ultimately taken with the Hubble flow, and thus does not end up within the main host at present-day.

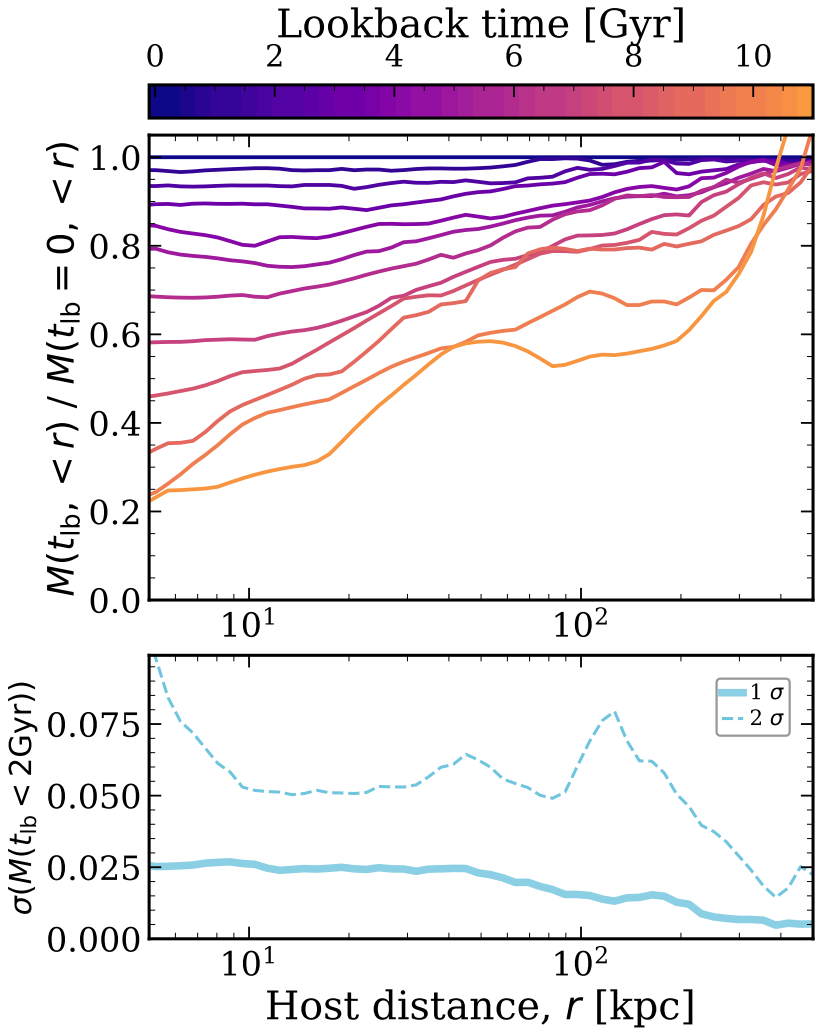

FIGURE 5.3. See the figure caption in the text.

Figure 5.3 caption: **Top**: Similar to Figure 5.2, the total mass within a given physical distance, r, relative to the value today, but now as a function of r. We show the median across our 13 MW/M31-mass hosts, at various lookback times, $t_{\rm lb}$, back to 11 Gyr, which encompasses the infall times of > 95 per cent of surviving satellites. The enclosed mass increases over time at essentially all r, with the inner halo experiencing the most fractional mass growth. Thus, the approximation of a static halo mass/potential is least accurate for satellites with the smallest pericenters. For context, the median recent pericenter distance of surviving satellites is $\approx 50-60$ kpc. **Bottom**: The 1- and

$2-\sigma$ standard deviation of the fractional change/fluctuation of the total mass of the host over the last 2 Gyr, relative to the average mass over that time. 2 Gyr is typical of the orbital timescale of satellites at $r \approx 29$ kpc, which is roughly half the typical recent or minimum pericenter distance. Both the 1- and $2-\sigma$ standard deviations tend to decreases with increasing r, so larger distances show smaller variations. On average, the host mass growth/fluctuations over the last 2 Gyr are $\lesssim 5$ per cent, though such fluctuations would be even higher for halos with a massive (LMC- or M33-mass) satellite.

In our final way of investigating the mass growth of the main host, we show the evolution of the enclosed mass profile across time from $r = 5-500$ kpc in the top panel of Figure 5.3, where r is the satellite distance from the MW-mass host. We first calculate the median profile over all 13 hosts at each distance and snapshot with ~ 1 Gyr time spacing; see the color-bar for the lookback time of a given mass ratio. Then we normalize the curves for each snapshot to the median profile of the hosts at present-day, $t_{\rm lb} = 0$ Gyr ago.

In general, at each r the enclosed mass ratio increases over time, and likewise, at each snapshot the enclosed mass ratio increases with distance. Similar to the results in Figure 5.2, for larger r, the enclosed mass ratios do not change as much as for smaller r. For example, typical recent pericenter distances for satellites in the simulations are $\sim 50-60$ kpc, and compared to present-day, the enclosed mass was only 74 per cent during typical satellite infall times. The median present-day distance of the satellite galaxies in the simulations is around 175 kpc, and at $t_{\rm lb} = 7.4$ Gyr ago, the enclosed mass was ~ 83 per cent of its $z = 0$ mass. However, near the virial radius and beyond, the enclosed mass was already 97 per cent during typical satellite infall times.

The enclosed mass at any given snapshot is subject to the perturbations of satellite galaxies that are orbiting around and merging within the main host. Thus, to better test the variability in the enclosed masses, we calculate a time-averaged enclosed mass profile over the last 2 Gyr, and normalize the enclosed mass at every snapshot between $t_{\rm lb} = 0-2$ Gyr to this time averaged one.

We then plot the 1 and 2σ values of these ratios at each distance in the solid and dashed blue lines, respectively, in the bottom panel of Figure 5.3.

Similar to the top panel, the 1σ curve shows larger variability at small r suggesting that the fractional growth in the inner regions of the MW-mass host is larger than at large r. However, in the 2σ curve we do not see a monotonically decreasing variability with distance. Instead, the 2σ variation is constant between $r = 10 - 100$ kpc. Between $r = 100 - 200$ kpc there is a bump in the 2σ curve, which might be associated with massive satellite galaxies within the last 2 Gyr.

We know that galaxies evolve and grow hierarchically over time, and each figure in this section explicitly quantifies this idea in the evolving virial region (Figure 5.1), within fixed distance apertures (Figure 5.2), and at fixed time (Figure 5.3). By modeling the enclosed mass, or potential, of a galaxy at $z = 0$, and holding the potential fixed across many Gyr, we do not account for the real and substantial growth of the environment in which smaller satellite galaxies orbit. We next investigate the common assumption that total orbital energy and specific angular momentum are conserved across time.

5.4.2. Satellite Energy and Angular Momentum. The total orbital energy of a satellite as it orbits around its MW-mass host is often expected to be a conserved quantity. The same is true for the specific orbital angular momentum of the orbit, $\ell = r \times v$, however, as we saw in Santistevan et al. (2023), this was only true for satellites that fell into their MW-mass host halos $\lesssim 10$ Gyr ago, and for satellites with $M_{\rm star} \lesssim 10^{7.5}\,{\rm M}_\odot$. Furthermore, the width of the 68th percentile and full distributions showed satellites either losing angular momentum since infall, or increasing ℓ by a factor of $\sim 2\times$ or more. We examine the extent to which the orbital energy of a satellite, and its angular momentum relative to present-day, are conserved in this section. However, we stress that we are showing trends across the full sample, and this are not meant to be interpreted on a satellite-by-satellite basis, but rather to glean insight in populations of satellites.

We show the difference in the total orbital energy between present-day and infall into MW-mass host in the top row of Figure 5.4. To avoid biasing the results from a single host, we additionally

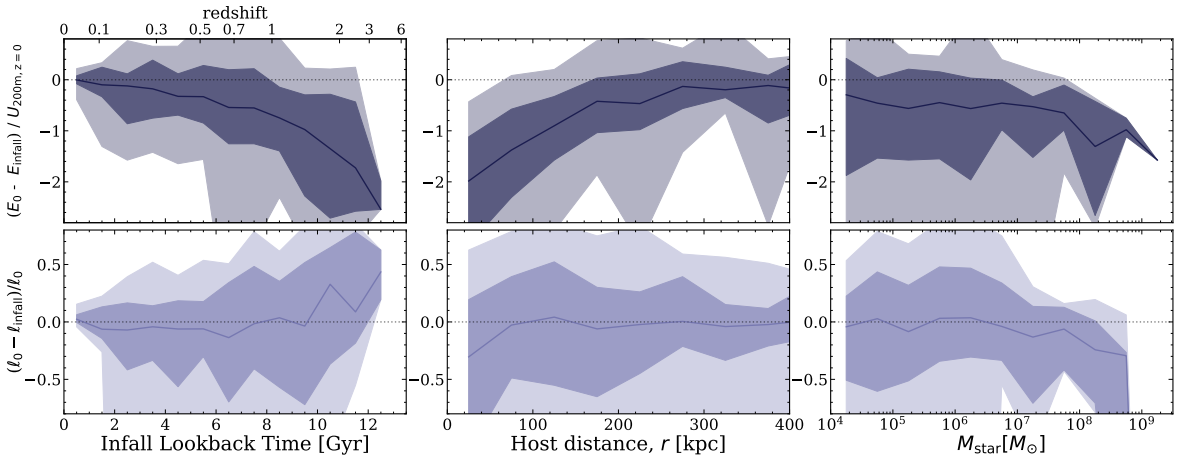

FIGURE 5.4. The difference in a satellite's orbital energy and angular momentum between today and at the time of first infall into the MW-mass halo, versus lookback time of infall (left), current distance from host, r (middle), and satellite $M_{\rm star}$ (right). Orbit modeling nearly always assumes that these quantities are conserved. Solid lines show the median for all satellites within the 13 MW-mass hosts, and the dark and light shaded regions show the 68th and 95th percentiles. **Top row**: Change in specific total energy, E, relative to the virial energy of the MW-mass host halo today, $U_{\rm 200m,z=0} = GM_{\rm 200m}/R_{\rm 200m}$. The median decreases from 0 to -2.5 with increasing infall lookback time, meaning that satellites have lost energy and become more bound since infall, because the host has grown (Figures 5.1-5.3). Because r inversely correlates with infall time, the median change in energy increases from -2 to -0.2 with r (middle), and is constant with satellite mass at $M_{\rm star} \lesssim 10^{7.5}\,{\rm M}_\odot$ but decreases at higher mass because of dynamical friction. Although E correlates with these properties, the mean widths of the 68th percentiles span ~ 1.24 versus infall time, 1.21 versus r, and 1.5 versus $M_{\rm star}$ highlighting the uncertainty for any given satellite. **Bottom row**: Change in specific angular momentum, ℓ, relative to the value today. The median fractional difference is generally zero for satellites that fell in $\lesssim 9.5\,{\rm Gyr}$ ago, but earlier infalling satellites have increased by up to ~ 45 per cent. We find little to no change in the median ℓ with r or $M_{\rm star}$, except for satellites with $M_{\rm star} \gtrsim 10^7\,{\rm M}_\odot$, which experienced strong dynamical friction. The mean widths of the 68th percentiles in the fractional change in ℓ span 70 per cent versus infall time and r, and 100 per cent versus $M_{\rm star}$. Thus, the uncertainties in both E and ℓ suggest that we cannot determine its energy or angular momentum from this plot to within a factor of $\gtrsim 2$ or so.

divide these energy differences by $U_{\rm 200m,0}$ which we define to be $G \times M_{\rm 200m}/R_{\rm 200m}$, the analytic gravitational potential at present-day for the mass enclosed within the virial radius, $R_{\rm 200m}$. We plot these fractional energy differences as a function of the lookback time of satellite infall into

MW-mass host, $t^{\rm lb}_{\rm infall,MW}$, distance from the MW-mass host, r, and satellite stellar mass, $M_{\rm star}$. We show the median trends across the full sample of satellites with the solid lines, and the 68th and 95th percentiles as the dark and light shaded regions, respectively.

Generally, the median fractional difference in total energy decreases with increasing $t^{\rm lb}_{\rm infall,MW}$ from 0 to −2.5 (top left). Thus, satellites that fell into their MW-mass hosts earlier have since lost more energy than recently infalling satellites. Satellites with $t^{\rm lb}_{\rm infall,MW} \lesssim 4\,{\rm Gyr}$ have likely just fallen into the MW-mass host which is why the median fractional energy difference is close to zero. The fractional energy difference then decreases significantly for earlier infalling satellites because their MW-mass host has grown in mass by $\gtrsim$ 20 per cent (Figure 5.1), and thus the potential well has deepened. However, at nearly all infall times, the width of the percentiles span larger than 100 per cent, which means the energy at infall cannot be approximated as the energy today to within a factor of ~ 2. Some satellites still have zero total energy with $t^{\rm lb}_{\rm infall,MW} \sim 11\,{\rm Gyr}$ ago, highlighting the diversity in the orbits of satellites.

Next, the fractional difference in total energy increases with r (top middle), from −2 to roughly −0.2 for satellites near $R_{\rm 200m}$. In Santistevan et al. (2023), we saw that $t^{\rm lb}_{\rm infall,MW}$ strongly correlates with r. Although the scatter in $t^{\rm lb}_{\rm infall,MW}$ is large, satellites that fell into their MW-mass hosts early typically orbit at smaller distances compared to satellites that fell in recently which orbit at larger distances. We see similar trends here, where satellites that currently orbit at smaller distances have lost more energy since their infall, compared to satellites that orbit at larger distances. This is likely because they fell in early, and thus the MW-mass host has grown more during the time since infall and the potential well is much deeper today. Satellites at large r only recently fell in, therefore their energy at present-day is comparable to their energies at infall.

The fractional difference in total energy only weakly depends on $M_{\rm star}$ (top right). Satellites below $M_{\rm star} \lesssim 10^{7.5}\,{\rm M}_\odot$ have a median fractional energy difference of roughly −0.5, and the fractional energy difference for more massive satellites decreases to as low as −1.6. These massive satellites feel stronger dynamical friction forces which cause them to slow down and ultimately

merge away quickly, even though more massive satellites tend to fall into their MW-mass host at later times. We finally caution the reader that there is only one satellite with $M_{\rm star} > 10^9\,{\rm M}_\odot$.

Next, we investigate trends in the specific angular momenta of satellites, ℓ. In Santistevan et al. (2023), we discussed trends in the angular momentum difference today versus infall, normalized by the angular momentum at satellite infall; $(\ell_0 - \ell_{\rm infall})/\ell_{\rm infall}$. We show the same difference, only now normalized by the angular momentum at present-day, i.e. $(\ell_0 - \ell_{\rm infall})/\ell_0$, in the bottom row of Figure 5.4.

Similar to Santistevan et al. (2023), we find weak dependence in ℓ since infall. The median present-day ℓ for satellites that fell in $\gtrsim 9.5\,{\rm Gyr}$ ago is as large as ~ 45 per cent compared to the values at infall. These satellites make up roughly 20 per cent of the total sample. Because these satellites fell into the MW-mass host early, they orbit at smaller distances, therefore this increase in ℓ may come from an increase in their total velocity over time. Satellites that fell in this early typically completed 4 pericenters since infall, so these satellites may also temporarily be at pericenter today where they orbit with faster velocities.

The fractional difference in ℓ shows virtually no dependence with r or $M_{\rm star}$. The average widths in each of the 68th percentiles is ~ 80 and 100 per cent versus r and $M_{\rm star}$, respectively. In short, we find qualitative agreement with Santistevan et al. (2023).

The trends we present here suggest that satellite *populations* do not show global energy conservation. If we focus on the 68th and 95th percentile ranges in each of the panels, we do see cases in which some satellites have total energies at present-day comparable to at infall, for a large range of infall time, r, and $M_{\rm star}$. However, this is certainly not true for *all* satellites, and the uncertainties in E and ℓ suggest that one cannot determine the satellite energy or angular momentum at infall to within a factor of $\gtrsim 2$. Rather it is more common for satellites to lose some of their orbital energy since first falling into their MW-mass host. This lack of energy conservation, as we discussed above, is likely due to both the growth of the MW-mass host potential over time, the effects of

dynamical friction on high-mass satellites, but also satellite-satellite interactions that may torque the satellite orbits.

We also investigated the extent to which the changes in the kinetic and potential energy components were conserved with infall time, r, and $M_{\rm star}$. Versus infall time, the fractional change in the kinetic energy of satellites that fell in recently was positive and the fractional change in the potential energy was negative, suggesting these satellites were likely on their first infall and nearing their first pericenter. Satellites with larger infall times often had negative fractional changes in kinetic and potential energies, due to both dynamical friction slowing the satellites down and the growing host potential over time. Within $r < 100\,{\rm kpc}$, the circular velocity profile rises significantly, and satellites that orbit today had much larger kinetic energies compared to at infall, and orbit in a deeper region of the host potential. We note no strong trends with $M_{\rm star}$. Finally, we checked the fractional change in E relative to their energy at infall, i.e. $(E_0 - E_{\rm infall})/E_{\rm infall}$. The trends relative to infall were qualitatively similar to the results in the top row of Figure 5.4, only the dynamic range in the fractional differences were much smaller, between 0.2 to -0.6 or so.

Figure 5.5 caption: Fractional change in specific orbital energy, E, (top) and specific orbital angular momentum, ℓ, (bottom), of satellite orbits versus lookback time. Solid line shows the median across all satellites and dashed line shows the width of the 68th percentile. We only include galaxies at a given lookback time that are satellites at that time, that is, after they fell into the MW-mass halo (including splashback satellites). **Top**: Total orbital energy. The median fractional change increases with lookback time and is ≈ 2 for the earliest-infalling satellites, that is, they were less bound than they are today, because the host has grown (Figures 5.1-5.3). The median of this fractional difference, which represents the systematic bias of orbit modeling across the whole population, reaches 25 per cent at 4.8 Gyr ago, 50 per cent at 7.9 Gyr ago, and 100 percent at 9.2 Gyr ago. Similarly, the width of the 68th percentile reaches 25 per cent at 0.7 Gyr ago, 50 per cent at 1.6 Gyr ago, and 100 per cent at 6.1 Gyr ago. **Bottom row**: Specific orbital angular momentum. The median fractional change is about zero over the last ≈ 6 Gyr, but at earlier times the

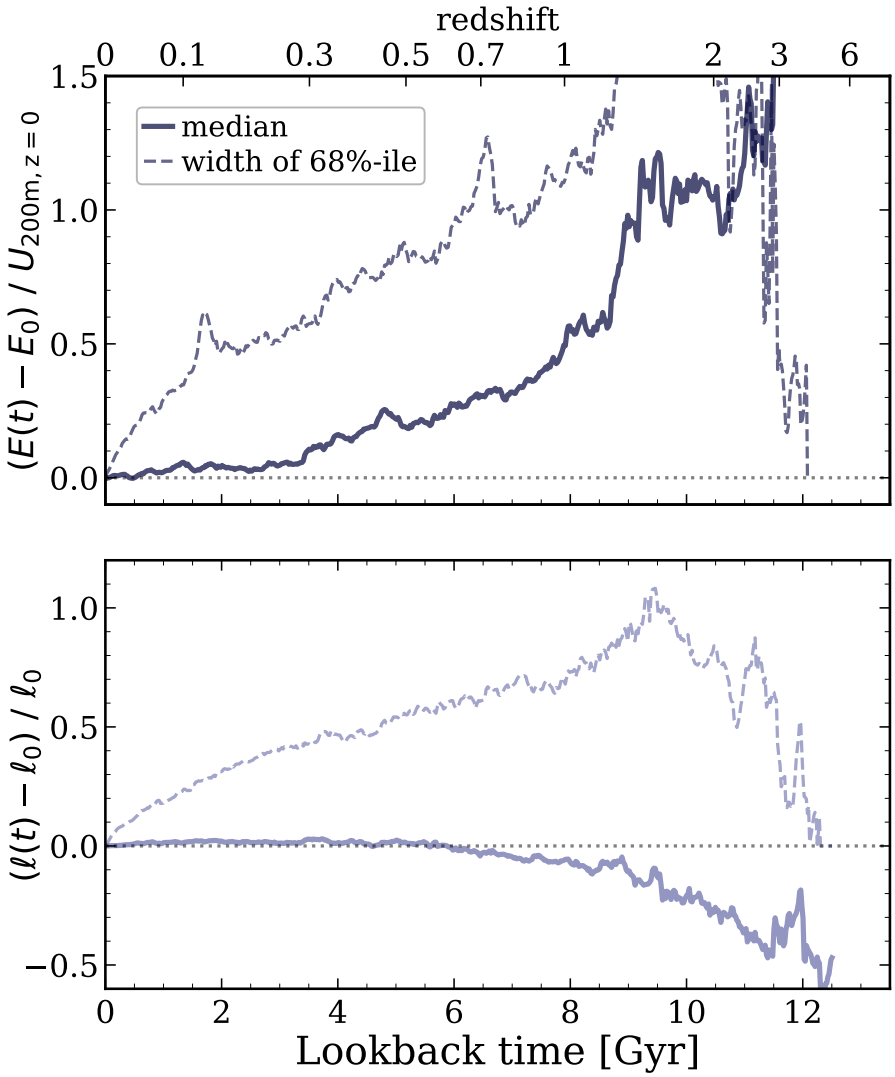

FIGURE 5.5. See the figure caption in the text.

angular momentum decreases to −60 per cent of the present-day value. Thus, the specific angular momentum of satellites generally increased over time because satellites fell into the MW-mass host halo at larger distances more recently, because the halo continues to grow. Because of the slower change in ℓ compared to E, the fractional change in ℓ reaches 25 and 50 per cent around 10.3 and 12.3 Gyr ago, respectively. The width of the 68th percentile, however, reaches these fractions

earlier than the median; 25 per cent at 1.5 Gyr ago, 50 per cent at 4.7 Gyr ago, and 100 per cent at 9.3 Gyr ago.

We now explicitly quantify the extent to which E and ℓ for the satellite populations are conserved over time in Figure 5.5. At each snapshot we include only the low-mass galaxies that have fallen into the host before that snapshot, including splashback galaxies. In other words, we only include the instantaneous satellite population at each snapshot, which means that the sample of satellites decreases with increasing lookback time. For this instantaneous satellite population, we show the median fractional change in E or ℓ in the thick solid lines, and the width of the 68th percentiles in the thin dashed lines, respectively.

In the top row of Figure 5.5 we show the difference in the total orbital energy of satellites between any given lookback time and present-day, $E(t) - E_0$, and normalize this by the MW-mass host gravitational potential at present-day, $U_{200\mathrm{m,z=0}}$. Over the last $\sim$ 3.5 Gyr, the median total energy is relatively unchanged, but at earlier times, this fractional difference increases with increasing lookback time from 0 to as large as $\sim$ 2 at 11.75 Gyr ago. The fractional change in E reaches 25 per cent at $\sim$ 4.8 Gyr ago, 50 per cent at 7.9 Gyr ago, and 100 per cent at 9.2 Gyr ago.

Although we see little evolution in the median fractional energy change over the last 3.5 Gyr ago, this does not apply to all satellites. In fact, 3.5 Gyr ago, the width of the 68th percentile spanned $\sim$ 65 per cent or so. We show that the width increases with increasing lookback time, and reaches 25 per cent at $\sim$ 0.7 Gyr ago, 50 per cent at 1.6 Gyr ago, and 100 per cent 6.1 Gyr ago. Because the width increases with time, deriving the total energy change for a given satellite becomes more uncertain. Only 28 per cent of the sample were satellites $\geq$ 9 Gyr ago, thus, because the number of satellites decreases, the width of the 68th percentile also decreases to 0 at 12 Gyr ago where there was only one satellite at that time.

These results agree with Figures 5.1-5.3. We showed that the mass of the MW-mass host increases with time at nearly all radii. As these satellites orbited within the MW-mass host halo, the mass, and therefore the gravitational potential of the host, grew over time, which affect the

satellite's total orbital energy. In the top left panel of Figure 5.4, the total orbital energy of satellites at the time of their infall was larger (more positive) than their energies at present-day, and satellites that fell in early often have much more negative energies today than at infall. A second-order effect from dynamical friction can cause some of the more massive satellites, $M_{\rm star} \gtrsim 10^7\,{\rm M}_\odot$, to lose their orbital energy over time as well, however, these more massive satellites tend to have fallen into the MW-mass host at later times.

We show the fractional change in the specific angular momentum of satellites, ℓ, in the bottom row of Figure 5.5. Similar to energy, we take the difference of ℓ at each snapshot with the value at present-day, only here we normalize the difference by the present-day ℓ. Unlike E, the median ℓ is conserved for much longer, over the last ~ 6 Gyr. At earlier lookback times, the median fractional difference in ℓ then decreases. The median fractional difference reaches 25 per cent at a much later time of ~ 11.9 Gyr ago, and 50 per cent at 11.4 Gyr ago. Compared to the fractional change in E, the width of the 68th per cent reach a given fraction later, 25 per cent at ~ 1.5 Gyr ago, 50 per cent at 4.7 Gyr ago, and 100 per cent at 9.3 Gyr ago.

These results strongly agree with those in the bottom left panel of Figure 5.4, where the median fractional difference in ℓ (between present-day and the time of infall) for satellites that fell into the MW-mass host $\lesssim 9.5$ Gyr ago was close to 0. Satellites that fell into the MW-mass host $\gtrsim 9.5$ Gyr ago, have a median ℓ today larger than at infall, which corresponds to negative values here in the bottom panel of Figure 5.5.

The results in Figures 5.4 and 5.5 show that neither E nor ℓ are conserved across all time for the satellite populations. The median E or ℓ may be conserved for a short amount of time, near present-day, however, the range of possible values for each satellite can, and does, vary up to a factor of 2 or more. The range of lookback times where the median E or ℓ remain relatively constant are between $0 - 3.5$ Gyr ago and $0 - 6$ Gyr ago, respectively, however, one should not assume that this applies to *each satellite*. Instead, one cannot claim that the energies or angular momenta are within a factor of 2 of present-day beyond 6.1 Gyr ago and 9.3 Gyr ago, respectively.

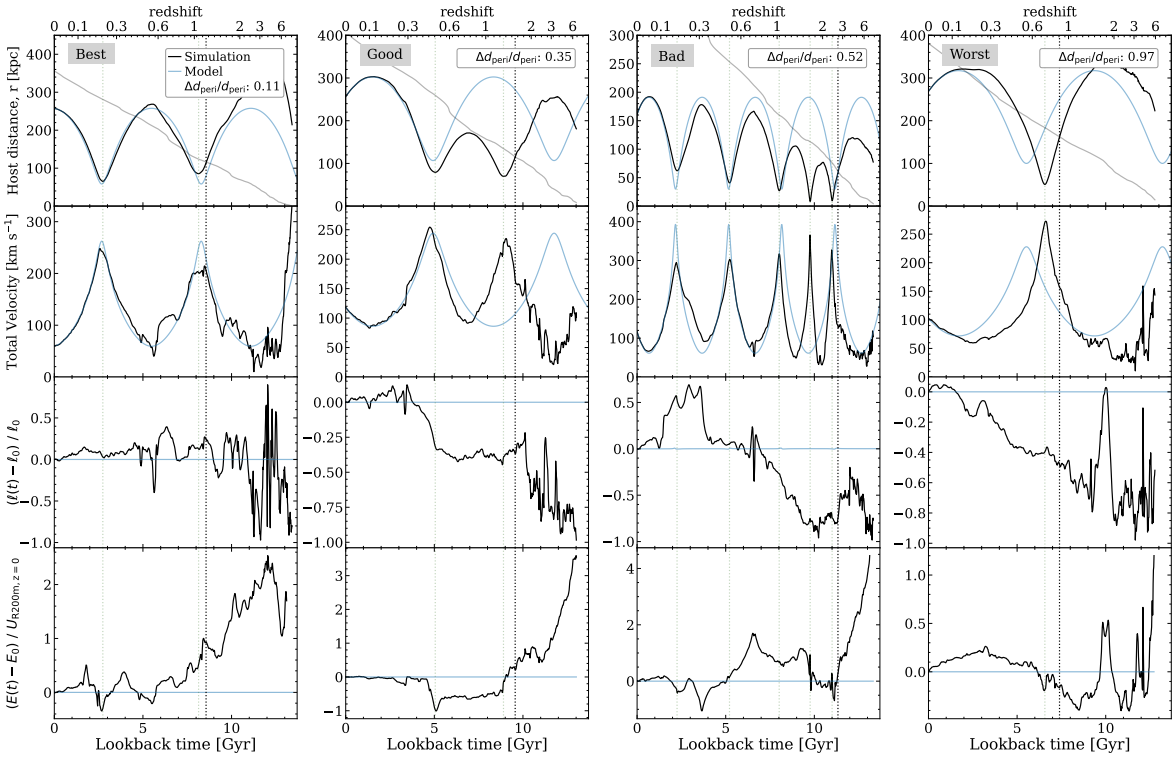

FIGURE 5.6. **Four case studies**. Orbital distance from the host galaxy, r (top row), total velocity (second row), specific angular momentum (third row), and specific total energy (bottom row). We show four satellites based on how well the most recent pericenter agrees between the simulation and model, $(d_{\rm peri,model} - d_{\rm peri,sim})/d_{\rm peri,sim}$, with increasing error from left to right (see legend). Black lines show values from the simulation, while blue lines show model-based orbits in a static axisymmetric host potential. The black vertical lines show when the satellites first fell into its MW-mass halo in the simulation, and in the top row the grey line shows the evolution of $R_{\rm 200m}$ of the host halo. The light green vertical dotted lines show when pericenters occurred in the simulation. In the left example, the orbit model and simulation agree for nearly two full orbits, while the right example shows agreement for less than half an orbit. Orbit modeling tends to recover the timing of the most recent pericenters better than their distances. As the bottom two rows show, specific angular momentum and total energy of the orbit are not generally conserved quantities; see Figures 5.4 and 5.5.

5.4.3. Static Potential Orbit Modeling. After discussing and quantifying the realistic growth of the MW-mass host in Section 5.4.1, and the extent to which energy and angular momentum are orbital invariants in Section 5.4.2, we now investigate the consequences of keeping the host mass, and potential, fixed over time. Here, we directly compare the satellite infall and orbit properties listed

in Section 5.3.4 between the simulations and the results we derived from numerically integrating satellite orbits within the static potential model outlined in Appendix B.1.

Figure 5.6 shows properties for four representative satellites in each column with varying degrees of how well the model reproduces the most recent pericenter distance. The legend in each column shows the value of $\Delta d_{\rm peri}/d_{\rm peri}$ for the most *recent* pericenter which we defined as $(d_{\rm peri,model} - d_{\rm peri,sim})/d_{\rm peri,sim}$. From left to right, we select a satellite within $\Delta d_{\rm peri}/d_{\rm peri} = [0, 0.25]$, $\Delta d_{\rm peri}/d_{\rm peri} = [0.25, 0.5]$, $\Delta d_{\rm peri}/d_{\rm peri} = [0.5, 0.75]$, and $\Delta d_{\rm peri}/d_{\rm peri} > 0.75$. The fraction of satellites within these selection windows is 73 per cent, 15 per cent, 6 per cent, and 6 per cent, respectively. The infall times of satellites for each selection window skew to more recent infall times at small $\Delta d_{\rm peri}/d_{\rm peri}$, and larger infall times at large $\Delta d_{\rm peri}/d_{\rm peri}$. For each property, we compare the simulation values in black to the model in blue. From top to bottom, we compare the total host distance (top row), total orbital velocity (second row), specific angular momentum (third row), and the total orbital energy of the satellites (bottom row). Finally, we also represent the time at which the satellite first fell into its MW-mass host in the vertical black dotted line, and the time at each pericentric passage in the vertical green dotted lines.

Because these are only four out of the 493 satellites in our sample, we refrain from an exhaustive comparison between simulation and model for these specific satellites, and rather discuss general qualitative differences. The first general trend we can see in both the host distance and total velocity is that the model agrees well with the simulations for a certain amount of time before diverging. In the left two columns the model recovers the orbits well for one half to two full orbits, the third column shows agreement with the timing of the orbit for two and a half orbits, but less-so with the distance and velocity, and in the right column the model matches the simulation for less than half an orbit. For the model, we do not account for the growth of the main MW-mass host over time, or other sources of gravitational potential like massive satellites, therefore the model orbits and properties are smooth compared to the simulation values which show perturbations, especially at earlier lookback times.

Even in the left two cases where the model does fairly well at reproducing the most recent pericenter distance, previous pericenters are not modeled as accurately, especially the timing of the pericenters which continues to become more out of phase with time. There are even cases as in the right column, where although the timing of the most recent pericenter is within ~ 0.5 Gyr, the distance is off by nearly a factor of 2.

Finally, what is especially apparent between the simulation and model is the lack of perfect conservation in specific orbital angular momentum (third row) and total orbital energy (bottom row). For each satellite, we show the fractional change in ℓ compared to present-day, i.e. $(\ell(t) - \ell_0)/\ell_0$. Even after a satellite falls into its MW-mass host, a satellite can increase or decrease $\gtrsim 50$ per cent of its angular momentum over time. The lack of conservation in ℓ is a combination of complex processes including the growth of the MW-mass galaxy and DM halo, satellite-satellite interactions, mergers, and the non-symmetric potential. However, a detailed investigation into the relative strengths of these effects is outside the scope of this paper. In the bottom row, we calculate the fractional change in energy compared to present-day, normalized by the hosts virial energy at $z = 0$, i.e. $(E(t) - E_0)/U_{\rm R200m,0}$. Similar to the results in Figures 5.4-5.5, the specific total energy of a satellite decreases over time, primarily due to the growth of the MW-mass host potential.

We now compare and discuss differences in specific orbit properties in the following subsections.

5.4.3.1. *Virial infall time.* In this section, we investigate and compare two different ways of calculating satellite infall time in the model, $t_{\rm infall,lb,model}$. For the first, we calculate by saving the lookback time of when a satellite first fell into its MW-mass host halo, while accounting for the growth of the halo's virial radius over time. In this first infall metric, we divide the satellite's current galactocentric distance, r, by the evolving $R_{\rm 200m}(t_{\rm lb})$ values, and save the corresponding time when this ratio first crosses under 1. However, we do not explicitly model the growth of the MW-mass host which means the satellite orbits are still periodic. In the second method, we keep the virial radius of the host halo at $z = 0$ fixed over time, and calculate when a satellite first crosses this value by similarly dividing the orbit distance by $R_{\rm 200m}(t_{\rm lb} = 0)$, and finding the first time this

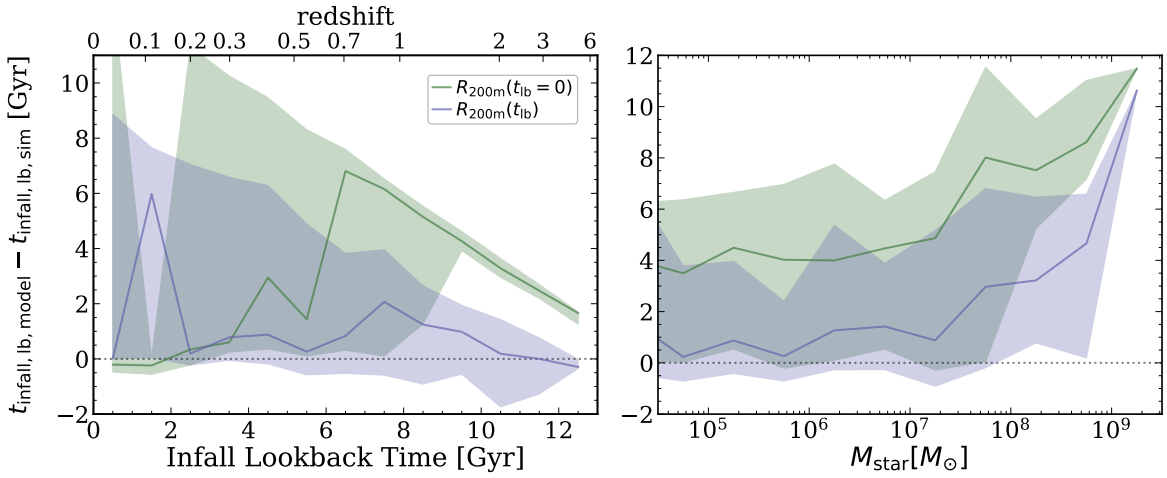

FIGURE 5.7. The difference in derived infall times from orbit modeling versus the simulation, as a function of infall time in the simulation (left) and satellite $M_{\rm star}$ (right). For orbit modeling, we measure infall times using both a non-evolving host halo radius, $R_{\rm 200m}(t_{\rm lb} = 0)$ (green) and using $R_{\rm 200m}(t_{\rm lb})$ from the simulation (purple). Solid lines show the median and shaded region shows the 68th per cent scatter across our satellites. **Left**: The median infall time is generally accurate for orbit modeling using an accurate $R_{\rm 200m}(t_{\rm lb})$, with a median difference of $\lesssim 2$ Gyr and an average scatter of 4.8 Gyr. The spike at 1.5 Gyr comes from orbit modeling over-predicting the infall times for most of satellites here by $\gtrsim 5$ Gyr, given errors in modeling recent apocenter distances. By contrast, orbit modeling using fixed $R_{\rm 200m}(t_{\rm lb} = 0)$ works well for recently infalling satellites but vastly overestimates the lookback time to infall for actual infall times $\gtrsim 4$ Gyr, with an average scatter of 5.6 Gyr. **Right**: The median and 68th percentile for both infall time metrics show similar trends with satellite mass, offset by ≈ 3 Gyr. The typical 68th percentiles span roughly 6.7 Gyr and 5.4 Gyr when using either the fixed $R_{\rm 200m}(t_{\rm lb} = 0)$ or accurate $R_{\rm 200m}(t_{\rm lb})$, respectively. Orbit modeling fails most significantly for satellites with $M_{\rm star} \gtrsim 10^7\,{\rm M}_\odot$, because the stronger dynamical friction for these satellites has shrunk their orbits since infall, and orbit modeling treats their orbits as small for all lookback times.

ratio crosses below 1. In this case, nearly 60 per cent of satellites in the model have always orbited within $R_{\rm 200m}(t_{\rm lb} = 0)$, so we define these model infall times to be 13.8 Gyr ago. We take the difference of each version of $t_{\rm infall,lb,model}$ to the infall time from the satellites in the simulations, $t_{\rm infall,lb,sim}$ (calculated with an evolving $R_{\rm 200m}$), and plot these differences in Figure 5.7 versus the infall times from the satellites in the simulations (left) and the satellite $M_{\rm star}$ (right). We show the

median infall times derived with an evolving virial radius, $R_{200m}(t_{lb})$ (purple), and with a fixed radius, $R_{200m}(t_{lb} = 0)$ (green), along with the 68th percentile range in the shaded regions.

In the left panel, we first investigate trends with satellite infall time. Focusing on trends with respect to the evolving virial radius, $R_{200m}(t_{lb})$, for nearly all infall times that we present, the median difference between the model and the simulation is generally within ~ 2 Gyr. The peak in the curve at 1.5 Gyr is driven by 11 of the 20 satellites in this particular bin in infall time, where the model predicted that these galaxies fell in $\gtrsim 5$ Gyr earlier than in the simulations. Similarly, a slightly smaller peak at 7.5 Gyr is caused by the model over-predicting $t_{infall,lb,model}$ for nearly half of the satellites by $\gtrsim 2$ Gyr.

The 68th percentile range is largest for the most recently infalling satellites, and decreases with increasing lookback time. The model generally overestimates the infall time compared to the simulation values because the model orbits are periodic and thus are more likely to cross the virial radius at earlier times. Across the full sample of satellites, the model over-predicts the infall time for roughly 65 per cent of them. Even when accounting for the evolving R_{200m}, the width of the 68th percentile spans $\gtrsim 2$ Gyr, which highlights the large uncertainty in the model.

When calculating infall time based on the fixed present-day $R_{200m}(t_{lb} = 0)$, the median trend shows relatively good agreement for satellites that fell in within the last ~ 4 Gyr. Beyond 4 Gyr ago, the median difference in infall time increases until ~ 7 Gyr ago where it decreases again. The orbits of these satellites were generally always within $R_{200m}(t_{lb} = 0)$ at all times, thus, the difference between the model and simulation infall times follows the relation $13.78\,\mathrm{Gyr} - t_{infall,lb,sim}$. We see a similar feature in the upper 68th percentile values. Of the subset of satellites that fell in between $1 - 2$ Gyr ago, only 2 of the 17 satellites were always within R_{200m}, which is why the 68th percentile dips down close to 0. However, it is clear that the associated uncertainties in this method of calculating infall time is much worse, and the width of the 68th percentile reaches as large as ~ 10 Gyr.

Versus $M_{\rm star}$, both infall metrics follow the same general trends in their medians: better agreement for satellites below $M_{\rm star} < 10^7\,{\rm M}_\odot$, and larger offsets for higher-mass satellites. The offset between the median and 68th percentile is roughly $3-4\,{\rm Gyr}$ between the two infall metrics. The values associated with the fixed $R_{\rm 200m}(t_{\rm lb}=0)$ method skew to larger values because many of the model orbits have always orbited within this distance Larger offsets at higher $M_{\rm star}$ are likely driven by dynamical friction which acts more efficiently at higher masses to rob the satellites of their orbital energy, and cause them to merge away relatively quickly; we do not account for this effect in the model. The widths of the 68th percentiles each span roughly $5.4-6.6\,{\rm Gyr}$, therefore, one cannot accurately determine a satellite galaxy's infall time at any given mass to within $\approx 5\,{\rm Gyr}$ or so.

Because the main idea of this paper is to compare derived results from our model to the simulations, throughout the rest of the paper, when we plot other properties versus infall lookback time, we are specifically plotting versus the infall times derived from the satellites in the simulations, which we calculate with an evolving $R_{\rm 200m}$.

5.4.3.2. *Pericenter properties.* We next investigate various properties associated with satellite pericenter events, which are the closest approaches to the host galaxy in a satellite's orbit. At pericenter, because a satellite is at its closest approach to the host, it thus orbits deeper in the host potential and experiences strong tidal disruption forces. It also orbits within a more dense region of the host CGM at pericenter (e.g. Samuel et al., 2022a), and therefore may experience strong ram pressure stripping in its gas.

Common static potential orbit modeling results in the literature suggest that satellite orbits only *shrink* over time because of dynamical friction and the time-dependent host potential (for example, see Weinberg, 1986; Taylor & Babul, 2001; Amorisco, 2017). This suggests that the most recent pericenter a satellite experienced should be the smallest, or minimum, that it has experienced. However, for many of the same satellites in our sample, Santistevan et al. (2023) showed that for 67 per cent of satellites with $N_{\rm peri} > 1$ the most recent pericenter is not the smallest, i.e. satellite

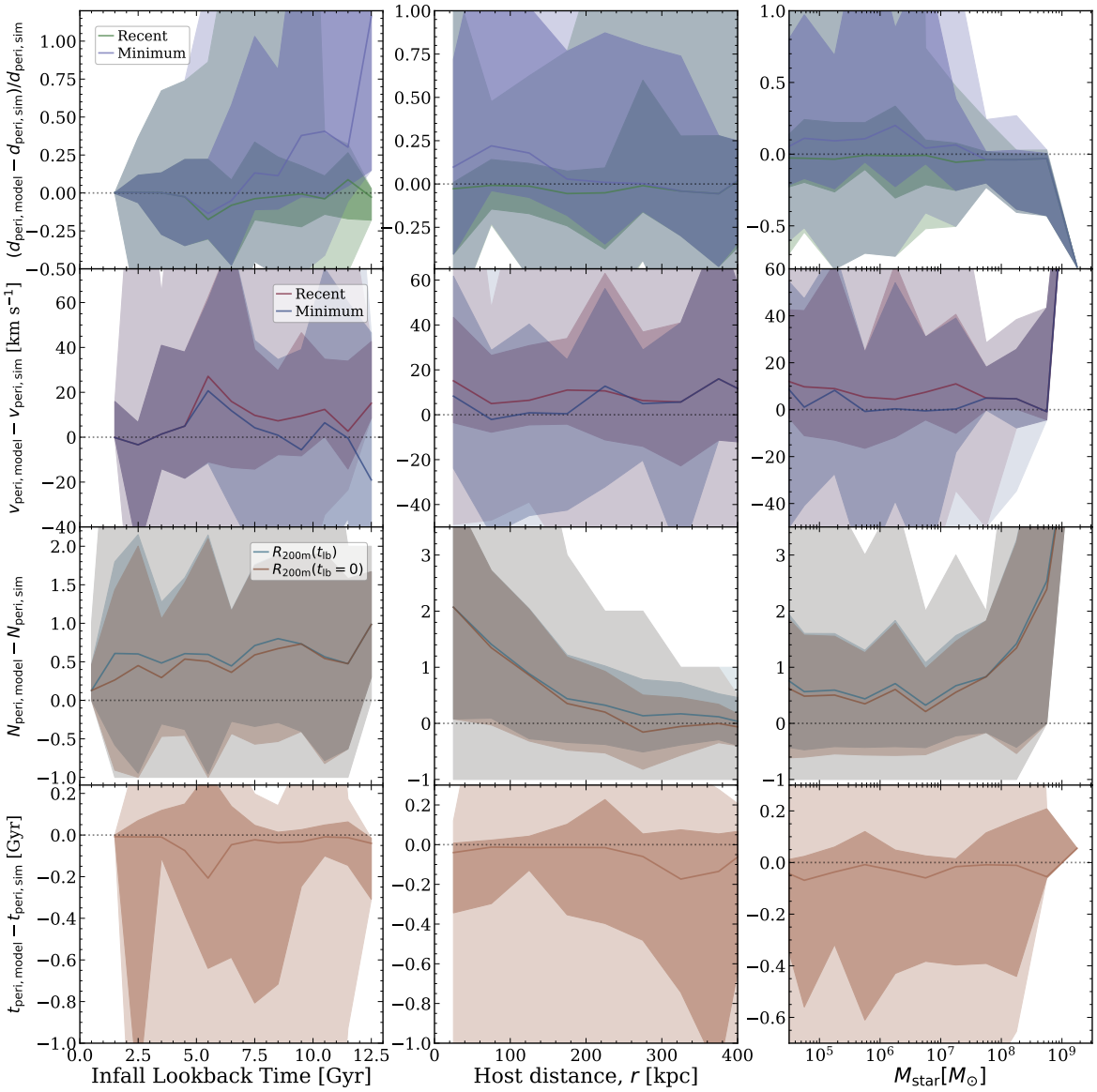

FIGURE 5.8. See the figure caption in the text.

orbits do not always shrink. For these reasons, in this section we present trends in both the most recent and minimum pericenter properties.

Figure 5.8 caption: A comparison of various properties of orbital pericenters, between the simulation and orbit modeling, for satellites at $z = 0$ versus their lookback time of infall into the MW-mass halo (left), distance from the MW-mass host, r (middle), and satellite M_{star} (right). Solid

lines show the median for satellites across all 13 hosts, and the dark and light shaded regions show the 68th and 95th percentiles, respectively. **First row**: The fractional difference between the most recent and minimum pericenter distances, $(d_{\rm peri,model} - d_{\rm peri,sim})/d_{\rm peri,sim}$. The model predicts larger minimum pericenters, as high as 100 per cent in the median, for satellites that fell in $\gtrsim 12\,{\rm Gyr}$ ago (left), and within 25 per cent for satellites at small r (middle), and for satellites $\lesssim 10^7\,{\rm M}_\odot$ (right). **Second row**: Difference between the total velocity, at the most recent and minimum pericenters, $v_{\rm peri,model} - v_{\rm model,sim}$. Orbit modeling generally over-predicts both the minimum and recent pericenter velocities, by $\approx 10-20\,{\rm km\,s^{-1}}$. **Third row**: Difference between the mean number of pericentric passages about the MW-mass host, since first crossing the growing host $R_{\rm 200m}(t_{\rm lb})$ and since first crossing $R_{\rm 200m}(t_{\rm lb} = 0)$. The difference in pericenter number slightly increases from $\approx 0-1$ with $t_{\rm lb}$ (left) and decreases from 2 to 0 with r (right), because satellites at close distance typically fell in earlier which means that they orbit longer in the model. **Fourth row**: Difference between the lookback times of the most recent pericenters, $t_{\rm peri,model} - t_{\rm peri,sim}$. The median is flat across all panels, and varies by at most $\approx 0.2\,{\rm Gyr}$ versus infall time (left). Although the median trends across the satellite populations agree well in most cases, the substantial widths in the percentiles for each property caution the considerable uncertainty associated with each.

In Figure 5.8, we show comparisons between the model and simulation for pericenter distances, $d_{\rm peri}$, the velocity at pericenter, v, the number of pericentric passages, $N_{\rm peri}$, and the timing of pericenters, $t_{\rm peri}$, as a function of lookback time of infall into the MW-mass host, present-day host distance, r, and satellite $M_{\rm star}$. For each pericenter property, $X_{\rm peri}$, we show the difference between the model and simulation with $X_{\rm peri,model} - X_{\rm peri,sim}$, except with pericenter distance, where we compare the *fractional* difference, with $(d_{\rm peri,model} - d_{\rm peri,sim})/d_{\rm peri,sim}$. In each panel, we show the median trends with the solid lines, and the 68th and 95th percentiles in the dark and light shaded regions, respectively.

We first compare trends in pericenter distance in the top row of panels in Figure 5.8. We compare both the most recent, $d_{\rm peri,rec}$, and minimum pericenter, $d_{\rm peri,min}$, in green and purple,

respectively. Because the orbits are periodic in the model, all pericenter distances are the same, however, we compare $d_{\rm peri,min}$ in the simulations to the pericenter in the model to highlight how the model may over/under-estimate this distance. With respect to satellite infall time (left panel), the model recovers the median $d_{\rm peri,rec}$ well. The median fractional difference is within $\sim$ 20 per cent, and the average width of the 68th percentile is roughly 37 per cent. The model recovers the median $d_{\rm peri,min}$ well for satellites that fell in $\lesssim$ 5 Gyr ago, but this is because all but one of these satellites experienced only one pericenter, which means the minimum is the most recent, and the two median lines (and 68th percentile ranges) are virtually identical. For satellites that fell in $\gtrsim$ 5 Gyr, the median fractional offset in $d_{\rm peri,min}$ diverges from the $d_{\rm peri,rec}$ trend and this offset between the two implies that the minimum pericenter is not the most recent, as Santistevan et al. (2023) noted. Roughly 60 per cent of satellites that fell in $\gtrsim$ 5 Gyr experienced multiple pericenters and because the model only predicts a single $d_{\rm peri}$ for a given satellite, these positive values suggest that the satellites in the simulations orbit at closer distances than in the model. The 68th percentile range for $d_{\rm peri,min}$ reaches 100 per cent around 7.5 Gyr ago, thus one cannot accurately predict the minimum pericenter to within a factor of $\gtrsim$ 2 for satellites that fell in earlier than this.

Next, the model and simulation median $d_{\rm peri,min}$ and $d_{\rm peri,rec}$ show general consistency across all satellite distances we present in the middle panel. The fractional difference in $d_{\rm peri,rec}$ is within $\lesssim$ 5 per cent, and within $\lesssim$ 25 per cent for $d_{\rm peri,min}$. As we will discuss below, $\sim$ 95 per cent of satellites that currently orbit beyond $r > 300$ kpc completed only one pericenter, thus, the median and 68th percentile range of values are the same between both $d_{\rm peri,min}$ and $d_{\rm peri,rec}$. Conversely, 2/3 of satellites currently within $r < 300$ kpc have $N_{\rm peri} > 1$ and similar to the left panel, the median and scatter for $d_{\rm peri,min}$ increases to primarily positive values indicating that the model is over-predicting for these satellites. For the satellites within 300 kpc, nearly 85 per cent of them fell into their MW-mass hosts over 5 Gyr ago. As with the left panel, the range in the 68th percentile is larger for $d_{\rm peri,min}$ than in $d_{\rm peri,rec}$, with average ranges spanning 111 per cent and 48 per cent, respectively.

Finally, the median fractional difference in both pericenter metrics show no dependence on $M_{\rm star} < 10^{8.75}\,{\rm M}_\odot$ (right panel). Similar to the other trends in pericenter distance here, the minimum pericenter is not equal to the most recent for low-mass satellites which typically fell in earlier, orbit at closer distances, and complete more pericenters. Only one satellite in our sample has a stellar mass above $M_{\rm star} > 10^9\,{\rm M}_\odot$, thus although the median drops significantly, we caution the reader in over-interpreting this feature.

In the second row of Figure 5.8 we investigate trends in the pericenter total velocity, $v_{\rm peri}$, and similar to $d_{\rm peri}$, we show trends in both the minimum, $v_{\rm peri,min}$ (lapis lazuli), and most recent, $v_{\rm peri,rec}$ (jam). Again, satellites at small infall time, large r, and high $M_{\rm star}$, only experienced one pericenter, so the median trends in both $v_{\rm peri,min}$ and $v_{\rm peri,rec}$ are the same. The model recovers both $v_{\rm peri,min}$ and $v_{\rm peri,rec}$ to within $\sim 30\,{\rm km\,s^{-1}}$ across all infall times and r, and nearly all $M_{\rm star}$. Although the model recovers the median $d_{\rm peri,rec}$ well, it over-estimates $v_{\rm peri,rec}$ in all panels presumably due to the lack of dynamical friction and gravitational tugs the satellites may feel from other satellites. Because the velocity at both pericenters is the same in the model, the fact that the model agrees more at $v_{\rm peri,min}$ indicates that the model does well in recovering the maximum satellite velocity a satellite has ever experienced.

Next, in the third row of Figure 5.8, we compare the number of pericentric passages a satellite experienced, $N_{\rm peri}$. Because the model orbits are periodic, it will predict as many pericenters a satellite can experience within the integration time, thus, we only count the number of pericentric passages a satellite experienced since infall into their MW-mass host. We count $N_{\rm peri}$ from the two infall metrics we discussed in Section 5.4.3.1: since infall into the MW-mass host while accounting for an evolving $R_{\rm 200m}$, labeled $R_{\rm 200m}(t_{\rm lb})$ (steel blue), and since infall while keeping $R_{\rm 200m}$ fixed at its present-day value, labeled $R_{\rm 200m}(t_{\rm lb} = 0)$ (tawny). In each of the panels, we show the *mean* and standard deviation for trends in $N_{\rm peri}$ because it is an integer property. Non-integer values in the difference in $N_{\rm peri}$ arise because this is a difference of two means, and need not be integer values.

In each of the three panels showing trends in $N_{\rm peri}$, the mean trends between the two different ways in which we calculate pericenter number differ by at most ~ 0.34. This indicates that it does not matter whether you model the growing $R_{\rm 200m}$ of the host galaxy, or simply use the value at $z = 0$. In the left panel, the mean difference in $N_{\rm peri}$ slightly increases with increasing infall time because the model orbits are periodic and there are no other forces acting on them except the global potential, so with longer integration times the model will predict larger $N_{\rm peri}$. The middle panel shows trends vs host distance, r, and we see that for satellites within $r \lesssim 250\,{\rm kpc}$, the model over-predicts $N_{\rm peri}$ because these satellites typically fell into their MW-mass hosts earlier than satellites at larger r. Finally, we see that the mean difference in $N_{\rm peri}$ is generally flat with $M_{\rm star}$ for satellites below $M_{\rm star} < 10^{7.5}\,{\rm M}_\odot$, but greatly increases for more massive satellites. Although high-mass satellites fell into their MW-mass hosts more recently, they do not experience many pericenters in the simulations because of the effects of dynamical friction and other gravitational tugs, while in the model, they are free to orbit as many times that fit into the integration time.

Interestingly, in about 8 per cent of satellites, the model predicts *fewer* pericentric passages than in the simulations. In these cases, the orbits near present-day in the simulations are at large distances, and because we use the present-day 6D phase space information as the 'initial conditions' for the model, the model therefore predicts large orbits with larger orbit periods.

The last pericenter property we compare is the timing of the recent pericenter, $t^{\rm lb}_{\rm peri}$. Because the orbits in the model are periodic, the timing of the minimum pericenter is undefined, therefore we only show the comparison in the most recent pericenter lookback time in the bottom row of Figure 5.8. The median difference is $\lesssim 0.2$ across all three panels. The median is also consistently negative, indicating that the model predicts the most recent pericenters to occur more recently than they do in the simulations. This is likely because in the model, the MW-mass host does not change in mass, and therefore is more massive than it should be at early times. Across the full sample, the model predicts later $t^{\rm lb}_{\rm peri,rec}$ than in the simulations for only 30 per cent of satellites.

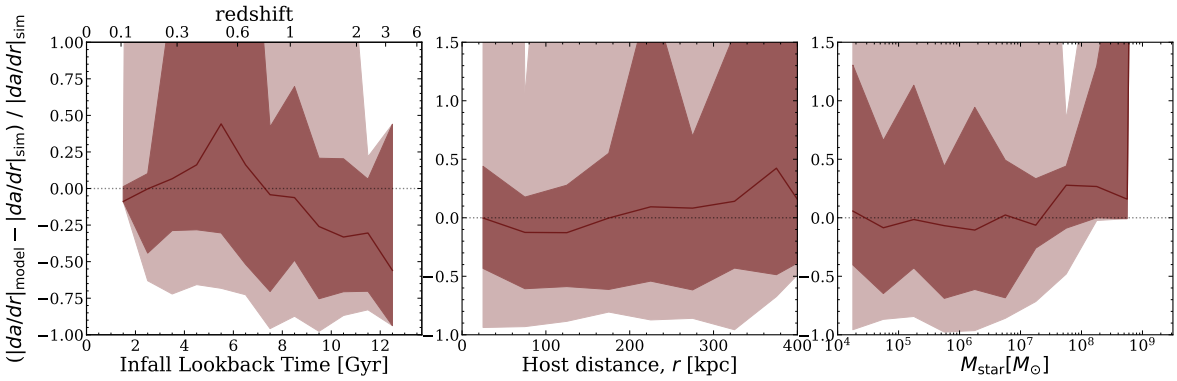

FIGURE 5.9. Comparing the maximum tidal acceleration, $|da/dr|$, from the MW-mass host that satellites experienced via orbit modeling versus in the simulations, as a function of lookback time of infall into the MW-mass halo (left), distance from the host, r (middle), and satellite $M_{\rm star}$ (right). Solid lines show the median across the satellite sample and the dark and light shaded regions show the 68th and 95th percentile ranges. These trends in the fractional difference in $|da/dr|$ mirror the trends in the minimum pericenter distance, $d_{\rm peri,min}$, in Figure 5.8 (top). Where the model under-predicts $d_{\rm peri,min}$, it over-predicts $|da/dr|$, and vice versa. Thus, for early-infalling satellites, because static orbit modeling does not account for the growth of the MW-mass host and the orbits are periodic, it increasingly over-predicts the pericenter value with increasing $t^{\rm lb}_{\rm infall,MW}$. Thus, the median $|da/dr|$ decreases with infall time for $t^{\rm lb}_{\rm infall,MW} \gtrsim 5.5$ Gyr. Satellites that fell in earlier orbited at smaller r and have lower $M_{\rm star}$, therefore the model over-predicts the minimum pericenter distance, and under-predicts $|da/dr|$ at small r and low $M_{\rm star}$. However, width of the 68th percentiles span more than 100 per cent in each of the panels here, so one cannot accurately infer $|da/dr|$ for a given satellite to $\lesssim$ a factor of 2.

As satellites orbit within the MW-mass environment, they feel tidal forces from the central galaxy which can either tidally disrupt them or torque their orbits. Here we present trends in the derivative of the tidal acceleration, which effectively measures the change in tidal force a satellite experiences across the satellite's diameter at any given position within the potential. We calculate the tidal acceleration to be:

$$a = \frac{G \times M(< r)}{r^2} \tag{5.5}$$

where $M(< r)$ is the enclosed mass within a distance r. We then take a numerical derivative of a with respect to r and save the maximum value of $|da/dr|$ a satellite experienced across all snapshots.

For the model, because the host mass/potential is fixed, we keep the present-day, $t_{\rm lb} = 0\,{\rm Gyr}$, mass profile fixed across time and calculate the maximum $|da/dr|$ value of the model orbits. We compare the trends in $|da/dr|$ between the model and simulation in Figure 5.9.

From Figure 5.8, we see that satellites the fell in between $t^{\rm lb}_{\rm infall,MW} = 2.5 - 7\,{\rm Gyr}$ ago typically have larger minimum pericenters in the simulations than in the model, which explains the positive bump in the fractional difference in $|da/dr|$ versus infall time in Figure 5.9 (left). The model predicts larger $|da/dr|$ values up to 45 per cent for these satellites. Satellites that fell in $t^{\rm lb}_{\rm infall,MW} \gtrsim 7\,{\rm Gyr}$ ago show larger minimum pericenters in the model, thus, the simulations predict larger values of $|da/dr|$ and the fractional difference decreases from ~ 0 to -55 per cent. The most recent infalling satellites however, $t^{\rm lb}_{\rm infall,MW} < 2.5\,{\rm Gyr}$ ago, have nearly equal pericenters in the simulation and model, which is why the median fractional difference and 68th percentile range of $|da/dr|$ is near zero. Because of this large uncertainty, one cannot recover values to within a factor of 2 for nearly any range of infall times.

Both the middle and right panels of Figure 5.9 show little to no dependence of the fractional difference in $|da/dr|$ versus r and $M_{\rm star}$. Although the median fractional difference is close to zero in both panels, the width of the 68th percentile increases from $0.78 - 2$ with r, and the mean width versus $M_{\rm star}$ is 145 per cent. In all three panels here, the width of the 95th percentile spans 100 per cent or more.

We additionally checked the values of $|da/dr|$ when the satellites were at their minimum pericenter, and these values differed from the maximum $|da/dr|$ a satellite experienced by $\lesssim 20$ per cent. Thus, because the maximum $|da/dr|$ correlates with the value of $|da/dr|$ at the minimum pericenter the trends in Figure 5.9 mirror the trends in $d_{\rm peri,min}$ in the top row of Figure 5.8.

5.4.3.3. *Apocenter, orbital period, and eccentricity. Figure 5.10 caption:* Comparing the most recent apocenter distance, $d_{\rm apo}$, orbital period, T, and orbital eccentricity, e, from orbit modeling versus in the simulations, as a function of lookback time of infall in the MW-mass halo (left), current distance from MW-mass host, r (middle), and satellite $M_{\rm star}$ (right). Solid lines show the median

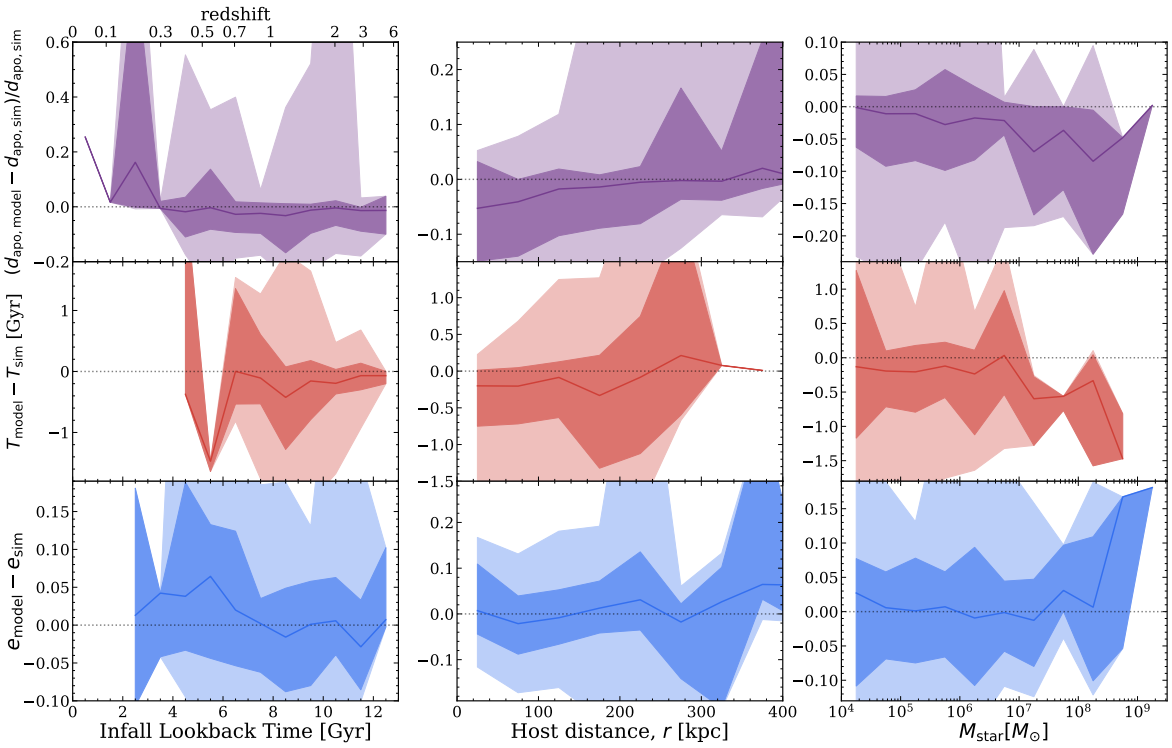

FIGURE 5.10. See the figure caption in the text.

and the dark and light shaded regions show the 68th and 95th percentiles. **Top row**: The most recent apocenter distance. Versus infall time, the median is roughly constant at −2 per cent for satellites that fell in $\gtrsim 3.5$ Gyr ago. The model recovers the median apocenter distance to within ±5 per cent versus r, but the median decreases slightly with $M_{\rm star}$ from ~ 0 to 8 per cent at $M_{\rm star} = 10^{8.25}\,{\rm M}_\odot$. Although the medians in each panel may only be within a few per cent, the widths of the 68th percentiles span $\approx 10-20$ per cent or more, underscoring the uncertainty for a given satellite. **Middle row**: Difference between the most recent orbital time, T, defined as the difference in time between the two most recent pericenters. Because the model generally under-predicts the recent pericenter lookback times, and because the orbits are periodic, the median difference in T is slightly negative across all panels, and even as low as $0.5-1.5$ Gyr for satellites with $M_{\rm star} \gtrsim 10^{7.25}\,{\rm M}_\odot$. **Bottom row**: Difference between most recent orbital eccentricity, $e = (d_{\rm apo} - d_{\rm peri})/(d_{\rm apo} + d_{\rm peri})$. The difference in e varies by at most 0.06 versus $t^{\rm lb}_{\rm infall,MW}$ and r, but satellites with $M_{\rm star} > 10^{8.5}\,{\rm M}_\odot$

have differences > 0.15. In general, the model recovers the median properties here within ~ 7 per cent (see Table 5.2), though with non-negligible 68th/95th percentiles spanning $12 - 29$ per cent, likely because these properties all depend on pericenter and apocenter events that occur in the recent past.

Although pericenters tell us about how far in the potential a satellite orbits, and allow us to approximate the tidal force a satellite experiences, they are only one extreme in the orbit. Apocenter events inform us how far a satellite galaxy orbits away from its host, and if it temporarily orbits outside of the virial radius, it will orbit through the hot, ionized CGM again which can further quench the satellite galaxy's star formation. Furthermore, a satellite will spend more time at apocenter because its velocity trends toward zero. In this section, and in Figure 5.10, we compare trends in the most recent apocenter distance, $d_{\rm apo,rec}$, the orbital period, T, and orbital eccentricity, e, versus the lookback time of infall into MW-mass host, $t^{\rm lb}_{\rm infall,MW}$, distance from the MW-mass host, r, and satellite $M_{\rm star}$.

In the top row of Figure 5.10, we compare trends in the most recent apocenter distance, $d_{\rm apo,rec}$. Of the 493 satellites in our sample, roughly $2/3$ of our sample (68 per cent) experienced an apocenter event. Similar to the way we calculate pericenter properties, we only count an apocenter event post-infall into the MW-mass host. Thus, the remaining 32 per cent of satellites either have yet to reach apocenter after their first pericenter or are currently on first infall. The most recent apocenter is also the maximum distance a satellite has ever experienced in approximately 29 per cent of our sample.

Versus infall time (top left), only two satellites that have experienced an apocenter in the simulation and model fell in $t^{\rm lb}_{\rm infall,MW} < 2\,{\rm Gyr}$ ago. 8 satellites fell into the host between $2 < t^{\rm lb}_{\rm infall,MW} < 3\,{\rm Gyr}$ ago, however, the model generally over-predicts the $d_{\rm apo,rec}$ by $\gtrsim 75$ per cent for half of them, which is why the median, 68th and 95th percentiles scatter to larger values. The fractional difference in apocenter distance is smaller for earlier infalling satellites, and the median is ~ 0.015 with a mean width of the 68th percentile of 0.12.

Figure 5.10 (top middle) shows very little dependence with r. Satellites that currently orbit at smaller distances generally fell into the MW-mass host earlier, thus, the model somewhat under-predicts $d_{\rm apo,rec}$ for satellites within $r \lesssim 250\,$kpc, similar to how it under-predicted $d_{\rm apo,rec}$ at large $t^{\rm lb}_{\rm infall,MW}$ (top left), however, these fractional differences are $\lesssim 5$ per cent. Overall, the mean width of the 68th percentile is ~ 0.15.

Finally, the median fractional difference in apocenter distance decreases weakly with $M_{\rm star}$ (top right). Only 3 satellites above $M_{\rm star} > 10^{8.5}\,{\rm M}_\odot$ that experienced an apocenter event. The lower mass satellites typically fell into their MW-mass host earlier, and they have smaller fractional differences in agreement to the top right panel versus $t^{\rm lb}_{\rm infall,MW}$.

The middle row of Figure 5.10 now shows trends in the most recent orbital period, T, of a satellite galaxy. We define the orbit period as being the difference between successive pericenter events, i.e. $t_{\rm peri,i+1} - t_{\rm peri,i}$, and in the figure, we show only the orbit period as calculated from the two most recent pericenter events. We also checked how these orbit periods compared to values we obtain by taking the difference of successive *apocenter* times, i.e. $T = t_{\rm apo,i+1} - t_{\rm apo,i}$, but we saw nearly identical results and there are more satellites with $N_{\rm peri} \geq 2$ than satellites with $N_{\rm apo} \geq 2$, so we choose the pericenter definition. Of the 493 satellites in our sample, only 47 per cent of them experienced 2 pericenters in both the simulation and the model.

In the middle left panel, satellites that fell in $t^{\rm lb}_{\rm infall,MW} < 4.5\,$Gyr ago did not have enough time to undergo two pericenters, and only 5 satellites fell in between $4.5 < t^{\rm lb}_{\rm infall,MW} < 6\,$Gyr ago with two pericenters. For earlier infalling satellites, the median difference in T varies by as much as $-0.4\,$Gyr, however, the mean across all infall times is smaller, -0.15. In the bottom left panel of Figure 5.8, we showed that the model under-predicted values for $t^{\rm lb}_{\rm peri,rec}$, thus, because T is defined from the two most recent pericenters, the median difference in T is negative across all infall times.

The difference in T is negligible versus r, and there are only two satellites beyond $r > 320\,$kpc that have experienced two pericenters (center panel). The model does not account for dynamical friction, therefore, for satellites with $N_{\rm peri} \geq 2$, if the model under-predicts $d_{\rm peri,rec}$ compared to the

simulations, it suggests more bound orbits and smaller T values, which is what we see for these satellites above $M_{\rm star} \gtrsim 10^{6.5}\,{\rm M}_\odot$.

Lastly, we compare the eccentricities of the satellite orbits in the bottom row of Figure 5.10. We calculate eccentricity with the following relation:

$$e = \frac{d_{\rm apo,rec} - d_{\rm peri,rec}}{d_{\rm apo,rec} + d_{\rm peri,rec}} \tag{5.6}$$

where $d_{\rm apo,rec}$ and $d_{\rm peri,rec}$ are the most recent pericenter and apocenter, respectively. In our sample, 65 per cent of satellites have eccentricity values in both the simulations and the model. We additionally compared the eccentricity values from the simulations to those returned by *galpy*, which calculates them in the same manner as Equation 5.6, and we saw identical values. For satellites that complete multiple pericenter and apocenter events, we also compared the eccentricity from the most recent events, to the average eccentricity across multiple orbits, and saw that the average was within ~ 0.2 of the most recent.

The bottom left panel in Figure 5.10 shows the trends in e versus $t^{\rm lb}_{\rm infall,MW}$. For satellites that fell in $t^{\rm lb}_{\rm infall,MW} \lesssim 4\,{\rm Gyr}$ ago, the model recovers $d_{\rm peri,rec}$ relatively well (Figure 5.8, top left), and the model over-predicts $d_{\rm apo,rec}$. Similarly, although the model recovers $d_{\rm apo,rec}$ well for satellites up to $t^{\rm lb}_{\rm infall,MW} = 7.5\,{\rm Gyr}$ ago, the model also under-predicts $d_{\rm peri,rec}$ for these satellites. These trends are what drive the difference in e to positive values. The model and simulation show similar results for earlier infalling satellites. The width of the 68th percentile varies by 0.29 at most.

The median difference in e is flat with r and $M_{\rm star} < 10^{8.25}\,{\rm M}_\odot$. Although the median and scatter increase to larger offsets with $M_{\rm star}$, the higher mass sample only make up about 2.5 per cent of satellites with eccentricities in the model and simulation.

All of the trends that we present in this section, and in Figure 5.10, rely on the most recent pericenter or apocenter properties. As we saw in Figure 5.8, the model does not do as well in recovering orbit properties which occurred earlier in time, such as properties related to the minimum pericenter a satellite experienced, consistent with the general results from D'Souza &

Bell (2022). Therefore, we expect subsequent orbit period and eccentricity calculations to show larger disagreements between the model and simulation.

5.4.3.4. *Recoverability of Orbit Properties.* In this paper, we compared 15 different properties of the satellites in the simulations to the model. Table 5.2 lists the properties we tested, as well as the median offsets and 68th/95th percentile widths of the distributions across our full sample of satellites. We compare both the raw differences of a given orbit property, X, between the model and simulation, defined as $X_{\rm model} - X_{\rm sim}$, as well as the fractional difference, defined as $(X_{\rm model} - X_{\rm sim})/X_{\rm sim}$. Additionally, we show the fractional change in the total energy since infall relative to the MW-mass potential, $(E_0 - E_{\rm inf})/U_{\rm 200m,0}$, as well as the fractional change in the specific angular momenta relative to present-day, $(\ell_0 - \ell_{\rm inf})/\ell_0$. We calculate these for only the simulated sample of satellites given that the total energy and angular momentum of a satellite is completely conserved in the axisymmetric model. To distinguish these measurements from the others, we include them below the horizontal line at the bottom of the table.

We describe the strength of the bias and uncertainty for a given satellite property in the two right-most columns of Table 5.2. Orbit properties are minimally, moderately, and highly biased if the median fractional offset between the model and simulation is between 0 – 10 per cent, 10 – 25 per cent, and > 25 per cent, respectively. Similarly, orbit properties are minimally, moderately, and highly uncertain if the width of the 68th percentile is between 0 – 25 per cent, 25 – 50 per cent, and > 50 per cent. Because the bias in a measurement is more important than the uncertainty, we impose more strict criteria in categorizing the strength of the bias, however, our choice in the exact binning for what constitutes minimal, moderate, and high bias/uncertainty is purely fiducial.

TABLE 5.2. See Table caption in the text.

	Raw Difference			Fractional Difference				
Property	Median offset	Width of 68%	Width of 95%	Median offset	Width of 68%	Width of 95%	Strength of Bias	Strength of Uncertainty
$d_{\mathrm{peri,rec}}$ [kpc]	-1.26	24.1	138	-0.025	0.42	2.38	Minimal	Moderate
$d_{\mathrm{peri,min}}$ [kpc]	3.23	37.0	160	0.066	1.07	7.11	Minimal	High
$t_{\mathrm{peri,rec}}$ [Gyr]	-0.03	0.50	3.43	-0.028	0.18	1.29	Minimal	Minimal
N_{peri}	0.51	1.00	4.00	0.25	0.49	1.94	Moderate	Moderate
$N_{\mathrm{peri,fixed}}$	0.53	1.00	4.00	0.26	0.49	1.944	High	Moderate
$v_{\mathrm{peri,rec}}$ [$\mathrm{km\,s^{-1}}$]	7.69	53.7	211	0.030	0.20	0.75	Minimal	Minimal
$v_{\mathrm{peri,min}}$ [$\mathrm{km\,s^{-1}}$]	3.10	86.9	264	0.012	0.33	0.86	Minimal	Moderate
$d_{\mathrm{apo,rec}}$ [kpc]	-2.75	24.2	195	-0.013	0.15	0.75	Minimal	Minimal
t_{inf} [Gyr]	-0.44	4.65	14.7	-0.067	0.82	5.34	Minimal	High
$t_{\mathrm{inf,fixed}}$ [Gyr]	4.17	7.10	17.6	0.44	1.09	5.84	High	High
e_{rec}	0.003	0.14	0.42	0.005	0.29	1.08	Minimal	Moderate
T_{rec} [Gyr]	-0.19	0.97	3.67	-0.071	0.27	0.82	Minimal	Moderate
$\lvert da/dr\rvert$ [$\mathrm{Gyr^{-2}}$]	-0.07	31.4	705	-0.017	1.32	8.40	Minimal	High
$(E_0 - E_{\mathrm{inf}})/U_{\mathrm{200m,0}}$	-	-	-	-0.537	1.63	3.65	High	High
$(\ell_0 - \ell_{\mathrm{inf}})/\ell_0$	-	-	-	-0.015	0.84	3.36	Minimal	High

Table 5.2 caption: Orbital history properties of satellite galaxies, comparing the results of orbit modeling in a static axisymmetric host potential against the simulations. Column list: property name; median offset, width of 68th percentile, width of 95th percentile. The left columns compare the raw difference, $X_{\rm model} - X_{\rm sim}$, while the middle columns compare the fractional difference, $(X_{\rm model} - X_{\rm sim}) \,/\, X_{\rm sim}$. Additionally, we describe the strength of the bias and uncertainty associated with each property. From top to bottom, the properties we present are: recent pericenter distance, $d_{\rm peri,rec}$, minimum pericenter distance, $d_{\rm peri,min}$, lookback time of recent pericenter, $t_{\rm peri,rec}$, number of pericenters since crossing MW-mass host $R_{\rm 200m}(t_{\rm lb})$, $N_{\rm peri}$, number of pericenters since crossing fixed MW-mass host $R_{\rm 200m}(t_{\rm lb} = 0)$, $N_{\rm peri,fixed}$, total velocity at recent pericenter, $v_{\rm peri,rec}$, total velocity at minimum pericenter, $v_{\rm peri,min}$, recent apocenter distance, $d_{\rm apo,rec}$, lookback time of infall into MW-mass host using $R_{\rm 200m}(t_{\rm lb})$, $t_{\rm inf}$, lookback time of infall into MW-mass host using fixed $R_{\rm 200m}(t_{\rm lb} = 0)$, $t_{\rm inf,fixed}$, recent orbit eccentricity, $e_{\rm rec}$, recent orbit period, $T_{\rm rec}$, maximum strength of the tidal field a satellite experienced, $|da/dr|$. In the bottom two rows, we show the difference in the total energy between present-day and the time of infall, normalized by the host potential energy at present-day, $(E_0 - E_{\rm infall})/U_{\rm 200m,0}$, and the fractional change in the specific angular momenta of satellites relative to present-day, $(\ell_0 - \ell_{\rm inf})/\ell_0$. Given that total energy and angular momentum are always conserved in the model, we place them below a horizontal line to distinguish them.

To better visualize which orbit properties are easier to recover in the idealistic model, Figure 5.11 shows the median offsets (left), and widths of the 68th and 95th percentiles (middle and left), for the fractional differences between the model and simulations; the points in this figure come from the columns in Table 5.2 under the 'Fractional Difference' header. In each panel, we rank-order the properties based on smallest to largest values. We additionally show a dotted horizontal line at 0 in the left panel to represent where the simulations and model agree. Finally, in this Figure, we shorten the labels for the fractional change in total energy and the fractional change in specific angular momentum simply as $E_{\rm inf}$ and $\ell_{\rm inf}$, respectively.

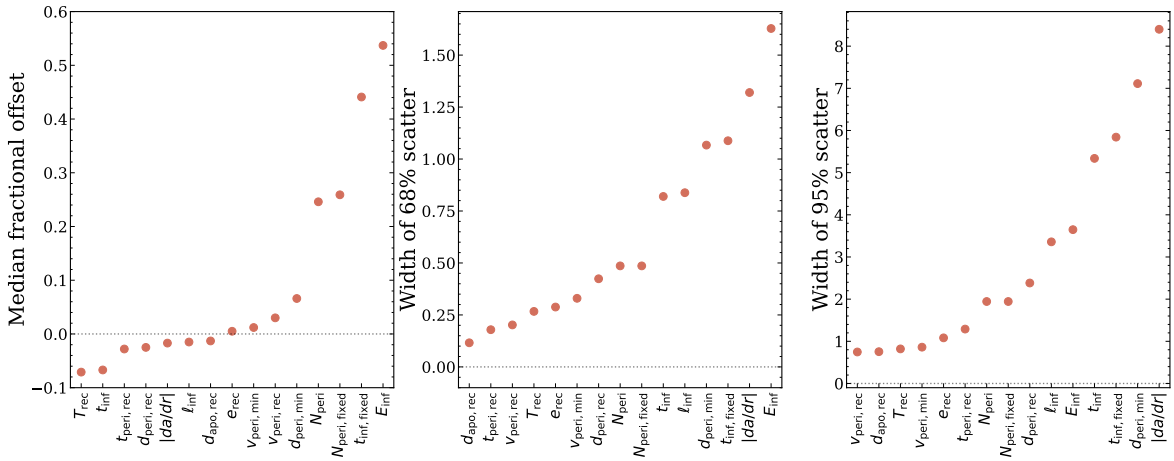

FIGURE 5.11. Rank ordering the 15 orbit history properties of satellite galaxies, as Table 5.2 lists, based on the fractional level of agreement between orbit modeling in a static axisymmetric host potential and the simulations. We shorten $(E_0 - E_{\rm inf})/U_{\rm 200m,0}$ to $E_{\rm inf}$, and $(\ell_0 - \ell_{\rm inf})/\ell_0$ to $\ell_{\rm inf}$. The left panel shows the median offset, in order from negative to positive values, and the middle and right panels show the widths of the 68th and 95th percentiles in order of agreement. Satellite orbit properties that occurred recently, such as the recent pericenter distances/times/velocities, show smaller offsets, $\lesssim 3$ per cent, compared to properties that occurred further in the past, such as the minimum pericenter distances or infall times, $\gtrsim 5$ per cent. The same is generally true for the width of the 68th percentiles, but not always for the 95th percentiles, despite the strong correlation between the two. Thus, deriving orbit parameters in the *recent* past using an idealistic model yields consistent results with the simulations, however, not including processes such as the host mass growth or satellite-satellite interactions will drive larger differences the longer you numerically integrate the orbits.

The median fractional offsets across all properties range between −0.54, for the fractional change in total energy, to 0.44, for the lookback time of satellite infall using a fixed $R_{\rm 200m}(t_{\rm lb} = 0)$. From left to right, the properties that show agreement to within ±5 per cent include the lookback time of the most recent pericenter, $t^{\rm lb}_{\rm peri,rec}$, the most recent pericenter distance, $d_{\rm peri,rec}$, the maximum value of the derivative of the tidal acceleration, $|da/dr|$, the fractional change in the angular momentum relative to present-day, $(\ell_0 - \ell_{\rm infall})/\ell_0$, the most recent apocenter distance, $d_{\rm apo,rec}$, the eccentricity of the most recent orbit, $e_{\rm rec}$, and the total satellite velocities at the minimum and most recent pericenters, $v_{\rm peri,min}$ and $v_{\rm peri,rec}$, respectively. Aside from $|da/dr|$, $\ell_{\rm inf}$, and $v_{\rm peri,min}$, these

properties that most agree between the model and simulation are *recent* properties. This suggests that deriving orbit properties that occurred more recently may be easier to recover.

The properties that show agreement within 5 – 10 per cent include the most recent satellite orbit period, $T_{\rm rec}$, the lookback time of infall into the MW-mass host, $t^{\rm lb}_{\rm infall,MW}$, and the distance of the minimum pericenter, $d_{\rm peri,min}$. Although we show the most recent orbit period, this is still defined as the difference in time between the two most recent pericenters, thus, requires properties slightly farther back in time. Similarly, both $t^{\rm lb}_{\rm infall,MW}$ and $d_{\rm peri,min}$ occurred farther back in time than the properties that best agree between the model and simulations.

The fractional offsets for the number of pericenters a satellite experienced when accounting for the evolving host $R_{\rm 200m}(t_{\rm lb})$, $N_{\rm peri}$, and when using a fixed $R_{\rm 200m}(t_{\rm lb} = 0)$, $N_{\rm peri,fixed}$ are comparable: $\sim$ 2.5 and 2.6 per cent, respectively. These properties strongly depend on the stellar mass of the satellite, its current distance from the host galaxy, or its infall time. Finally, the properties that agree least between the model and simulation are the lookback time of infall when keeping the virial radius of the host fixed, i.e. $R_{\rm 200m} = R_{\rm 200m}(t_{\rm lb} = 0)$, and the change in orbital energy since infall, $E_{\rm inf}$. The larger discrepancy here for infall time is because when we keep the virial radii fixed over time, some of the satellites in the model never orbit outside of this distance, thus we set their infall times to 13.78 Gyr, which is completely unrealistic.

The widths of the 68th percentiles range from $\approx 0.1 - 1.30$, and the widths of the 95th percentiles range from $\approx 0.8 - 8.5$. The relative ordering of properties with smaller percentile widths is not perfectly conserved, but the 6 properties with the smallest 68th percentile widths are the same 6 properties with the smallest 95th percentile widths. Similar to the median offsets, the properties that occurred more recently often have smaller 68th percentile widths (aside from $v_{\rm peri,min}$), but the same is not true for the 95th percentiles. Finally, we see no strong correlation between the median offsets and either percentile width.

5.5. Summary & Discussion

5.5.1. Summary of Results. We compared 15 orbit properties for 493 satellite galaxies around 13 MW-mass hosts in the FIRE-2 suite of cosmological simulations against a static, axisymmetric potential model to quantify the strengths and limitations of this orbit modeling technique. We model the potentials of the MW-mass hosts at $z = 0$ to within a few percent, therefore our orbit property comparisons represent lower-limits when using this technique.

We will now discuss the key questions we raised in the Introduction and our corresponding results.

1 How much has the MW-mass evolved during the time that these satellites first started orbiting in this environment and how does this affect our interpretation of results in static potential orbit modeling?

- During the times in which satellites first fell into their MW-mass hosts, 3.4 – 9.7 Gyr ago, the virial mass and radii of the host were between 33 – 86 per cent and 26 – 73 per cent of their present-day values (Figure 5.1). During the median infall time, the median virial mass and radius were 54 and 43 per cent of their values today, indicating that the host mass and radius have nearly doubled since these low-mass galaxies first became satellites.
- The total enclosed mass within a fixed distance increases with time, and similarly, the enclosed mass at any given time increases with distance (Figures 5.2-5.3). For instance, the enclosed mass within 50 kpc 10 Gyr ago, similar to the median recent pericenter distance, was only $\sim$ 60 per cent of the enclosed mass within 50 kpc at present-day.
- The fractional change in the enclosed mass at smaller distances is greater at small r than at large r (Figures 5.2-5.3). This is contrary to the 'inside-out' growth of the dark matter halo, where the inner regions form earlier than the outer regions of the halo.

2 Given our understanding of the orbital properties of satellite galaxies in the simulations, how well does the idealized model recover these properties, and are there properties that are easier to recover than others?

- The total orbital energy for a population of satellites is generally not conserved. Satellites that fell into their MW-mass host halos $\gtrsim 9.5\,\mathrm{Gyr}$ ago, satellites currently within $r \lesssim 120\,\mathrm{kpc}$, and satellites with $M_{\mathrm{star}} \gtrsim 10^8\,\mathrm{M}_\odot$ have typically lost energy, and their present-day energy is more negative by a factor of $\gtrsim 2$ compared to at infall (Figure 5.4, top row).
- The median specific angular momentum of satellites today is similar to the values at infall, however, the widths in the 68th percentiles reach as large as ~ 100 per cent versus infall time, r, and M_{star} (Figure 5.4, bottom row).
- Calculating the infall time of a satellite in the model with a growing $R_{200\mathrm{m}}(t)$ yields more consistent results with the simulations, $\lesssim 2\,\mathrm{Gyr}$ offset, compared to using the fixed $R_{200\mathrm{m}}(t = 0)$ at present-day, however, the 68th percentiles in both metrics can range as high as $\sim 10 - 12\,\mathrm{Gyr}$ (Figure 5.7).
- Orbit history properties that occurred more recently have smaller fractional offsets and uncertainties than properties that occurred in the past. For instance, the timing and distance of the most recent pericenter have median fractional offsets (uncertainties) $\lesssim 3\,(18-42)$ per cent, compared to the minimum pericenter distance, which occurred farther back in time, which has a fractional offset (uncertainty) of 7 (107) per cent (Figures 5.7 and 5.8 and Table 5.2).
- The orbit properties that are recovered best in the model (with smallest bias and uncertainty) include the timing and velocity of the recent pericenter, and the most recent apocenter distance. The properties that are recovered moderately well (moderate bias and/or uncertainty) include the recent pericenter distance and velocity, number of pericentric passages, and the most recent orbit eccentricities and orbit periods. The

properties that are not recovered well (largest bias and/or uncertainty) include the minimum pericenter distance, the lookback time of infall into MW-mass host halo, maximum strength of the tidal field, and the change in total orbital energy and angular momentum since infall (Figure 5.11 and Table 5.2).

- Even with near perfect knowledge of the mass distribution/potential at $z = 0$ in the host galaxies, the typical uncertainties in these orbit properties range from 15 per cent to > 100 per cent. Furthermore, the satellite-to-satellite variations in each, i.e. widths of the 95th percentile, are ≥ 75 per cent. The uncertainties suggest that one cannot recover orbit properties to within a factor of ~ 2 or more, and cautions against overgeneralizing the results for a single satellite from the median trends (Figure 5.11 and Table 5.2).

3 How far back in time can we reliably integrate the orbits of satellites?

- The total orbital energy of a satellite is not conserved across all time. Compared to present-day, the median total orbital energy is constant for the last ~ 3.5 Gyr, but reaches a 25, 50, and 100 per cent difference around 4.8, 7.9, and 9.2 Gyr ago. The uncertainty associated with a satellite's total energy reaches 25, 50, and 100 per cent around 0.7, 1.6, and 6.1 Gyr ago (Figure 5.5).
- The specific angular momentum of a satellite is also not conserved across time. The median difference is zero for the last ~ 6 Gyr ago, but reaches a difference of 25-50 per cent around $10.3 - 12.3$ Gyr ago. Similarly, the uncertainty associated with the angular momentum reaches 25-50 per cent around $1.5 - 4.7$ Gyr ago (Figure 5.5).

5.5.2. Discussion. Our analysis closely resembles the comprehensive analysis of D'Souza & Bell (2022), which similarly fit symmetric models to MW-mass hosts from the dark matter only versions of the ELVIS simulation suite (Garrison-Kimmel et al., 2014) to study the various uncertainties associated with this orbit modeling technique. Before we compare our results, it is important to note the non-trivial differences in methods between their study and ours to give more

context. First, D'Souza & Bell (2022) use dark matter only simulations, which neglect the strong tidal forces from the central galaxy that would otherwise disrupt any satellites that orbit nearby. The internal stellar feedback in a satellite also acts to reduce the inner dark matter densities and make the satellites more vulnerable to tidal disruption (Bullock & Boylan-Kolchin, 2017), therefore, tidal disruption is stronger in the baryonic simulations, albeit a second-order effect (Garrison-Kimmel et al., 2017b). Without these baryonic effects and processes, the satellites in dark matter only simulations typically fall into their MW-mass hosts earlier and are then able to complete more pericenters, while also orbiting *closer* to the center of the host with smaller pericentric passages as we have shown in Santistevan et al. (2023). However, D'Souza & Bell (2022) do account for the effects of dynamical friction which acts to slow satellites down and ultimately merge within the host, which we do not. D'Souza & Bell (2022) also include MW-mass hosts with LMC-like analogs. Because the LMC is so massive, it hosts its own satellite population, and studies suggest that it is on first infall into the MW and near its first pericenter (for example Kallivayalil et al., 2013; Deason et al., 2015; Kallivayalil et al., 2018; Patel et al., 2020), so accounting for its gravitational influence on the surrounding satellites is of great interest. Another significant difference between D'Souza & Bell (2022) and our analysis is that they account for the true mass growth of the MW-mass host at every snapshot by updating their potentials while keeping the potential fixed between snapshots.

D'Souza & Bell (2022) define a recovered property as being when the absolute value of the fractional difference is less than 30 per cent, i.e. $|X_{\rm true} - X_{\rm model}|/X_{\rm true} < 0.3$, and report the fraction of satellites that do not meet this criteria as being 'outliers'. They center their results on two hosts, 'iDouglas' which is an isolated MW-mass galaxy and 'iOates' which has an LMC analog, although they test orbits in iOates with and without the gravitational contribution from this massive companion. Similar to our analysis, they focus on the timing and distance of various pericenter events, the apocenter distances, and the infall times of satellites, and find that more recent pericenters and apocenters have smaller outlier fractions than pericenters or apocenters that happened at earlier times. In particular, the outlier fractions for the most recent, and second-most

recent pericenter distances in the hosts without the additional massive satellite are 31.2 – 47 and 43.8 – 69.9 per cent, respectively. Although we do not look specifically at the second-most recent pericenter, we do find that the median fractional offsets and 68th percentile widths for the most recent pericenter distance, −2.5 and 42 per cent, are smaller compared to the same for the minimum pericenter distance which often occurred ∼ 6 Gyr earlier, 0.066 and 107 per cent. The authors also show that the timing of the most recent pericenter is often better recovered than the distance, with an outlier fraction of 13.8 – 23.2 per cent. Our results show that the median fractional offset is −2.8, which is comparable to the offset in recent pericenter distance, but with a smaller 68th percentile width of 18 per cent. Thus, it is often easier to recover the timing of the pericentric event, rather than the distance associated with it.

In D'Souza & Bell (2022) they similarly show that the distance of the most recent apocenter has a smaller outlier fraction than the most recent pericenter distance, with a value of only 6.2 – 34.9 per cent. Our work also suggests that the apocenters are easier to recover, with a median fractional offset and 68th percentile width of −1.3 and 15 per cent, respectively. They conclude that apocenters are easier to model because they only depend on the binding energy of a satellite galaxy, while the pericenters depend on the angular momentum of a satellite as well as its binding energy. Properties at apocenter also do not intricately depend on the details of the gravitational potential at small distances like pericenters do, and rather, what is more important is modeling the total enclosed mass precisely, as both studies have done.

Finally, D'Souza & Bell (2022) also calculate the infall times of satellites and find good agreement in their simulations and model, with an outlier fraction of 11.2 – 28.9 per cent. Although we generally see small median offsets when calculating infall time with an evolving $R_{200\mathrm{m}}(t)$, ∼ 6.7 per cent, the associated uncertainty is high, ∼ 82 per cent, and using a fixed $R_{200\mathrm{m}}(t = 0)$ is worse. Without comparing to the model, Santistevan et al. (2023) showed that even though infall time strongly correlates with other properties, the full distribution of infall times for a given energy,

angular momentum, or distance can still span ~ 13 Gyr, indicating that the time of infall cannot be accurately approximated with present-day satellite properties.

D'Souza & Bell (2022) thoroughly explore other models for the growth of the MW-mass host, and most relevant to our work is their results which keep the total mass fixed over time, as we do here. They report that depending on the property that you are most interested in recovering, it can actually be a good approximation to use the *static* model. The outlier fractions for the most recent pericenter and apocenter distances in their static model are $28.8 - 37.3$ per cent and $10 - 13.3$ per cent, respectively, which are better than the model accounting for the true mass growth. However, previous events often have comparable or larger outlier fractions than the models that account for the true mass growth of the MW-mass host. The timing of the most recent pericenters have outlier fractions between $10 - 13.3$ per cent, which are also better than their fiducial model, but the infall times are not, with outlier fractions of $20 - 38.6$. Thus, as one integrates longer in time, the static model becomes less representative of the MW-mass environment, and thus satellite orbit properties that occurred earlier are not modeled as well, which is what we similarly see in the minimum pericenter properties and infall times. They do note that as long as there is a well motivated model of mass growth, the outlier fractions of events that occurred farther back in time become easier to recover again, even if you do not know the mass growth perfectly.

In their analysis, D'Souza & Bell (2022) investigate the various sources of uncertainty that may further reduce the accuracy of results when using static potential orbits modeling. The authors compare uncertainties in the virial mass of the MW-mass host, and thus uncertainties in the potential, the uncertainty in the 6D phase-space coordinates of the satellites at $z = 0$, the uncertainty in modeling the recent LMC-like accretion, and the uncertainty in modeling the motion of the MW-mass system as it moves throughout the Universe. They conclude that the uncertainties in the recovered orbit history properties when using simple parametric forms of the potential are comparable to the uncertainty in results caused by a ~ 30 per cent uncertainty in the virial mass of the MW-mass host. D'Souza & Bell (2021) mention that the uncertainties in their results are

also comparable to the uncertainties one gets when not accounting for the LMC-like companion. Without the LMC-like companion, the outlier fractions for the recent pericenter and apocenter distances increase from 36 and 23 per cent to 47 and 35 per cent, respectively, and the same is true for subsequent events, the timing of the pericenters, and the infall times.

To derive the orbit histories of satellites within the LG, many studies commonly use the static potential orbit modeling approximation, or simple approximations of the growth of the main galaxy. For instance, Kallivayalil et al. (2006) used a fixed MW-mass potential, as well as an LMC-like potential, to determine that the SMC was gravitationally bound to the LMC. In Kallivayalil et al. (2013) the authors used three epochs of HST measurements to constrain its proper motions and suggest that the LMC is likely on its first infall as previous studies have suggested (for example Besla et al., 2007). The authors compared different static models of the MW based on its virial mass, and models that account for the growth of the MW-mass halo over the last 10 Gyr, however, in the evolving models, the authors do not account for the central galaxy. Another work by Patel et al. (2017) similarly aimed to understand the orbit of the LMC around the MW, as well as the orbit of M33 around M31, and concluded that like the LMC, M33 is likely fell into the M31 halo less than $\lesssim 2$ Gyr ago or so. Although these authors modeled the many components of the main galaxy, they did not account for any time dependence of the mass or potential.

Other studies account for the gravitational contribution of the LMC on the other satellites of the MW. In Kallivayalil et al. (2018), the authors used Gaia data in conjunction with the dark matter only Aquarius simulation to determine which satellites might be gravitationally bound to the LMC. The simulation does have an LMC analogue, with a similar position and velocity at $z = 0$, however, does not account for the gravitational effect of the central galaxy. In a similar study by Patel et al. (2020), the authors use new Gaia data and numerically integrate the orbits of satellites in a model of the global potential with a MW, LMC, and SMC component to determine the orbital histories of the LMC and its satellites. Although they keep the potential fixed over time, they only backward integrate the orbits of satellites for ~ 6 Gyr, given that beyond this time,

MW-mass galaxies typically have $\lesssim 80$ per cent of their $z = 0$ mass. The authors conclude that the derived orbits for satellites of the LMC strongly depend on whether or not you include the LMC in the global potential, and in some cases, the contribution from the SMC is important as well. Finally, other studies aim to accurately model the LMC potential with basis function expansions fit to simulation data to understand how it affects the MW, and its satellites, as in Garavito-Camargo et al. (2019, 2021), however, doing so while accounting for the growth of the MW still remains challenging.

Studies such as Fritz et al. (2018a) and Fillingham et al. (2019) provide an exhaustive derivation of infall times, pericenter and apocenter distances, and orbit eccentricities for all satellites in the MW. Fritz et al. (2018a) uses data from Gaia DR2, and numerically integrate the orbits of all satellites in galpy in two different models for the MW potential, which they keep fixed over time. Depending on the mass of the MW, the authors show that some of the apocenters for the satellites can lie either inside or outside of the virial radius, which has implications for instance in studying how a satellite interacts with the hot gas in the MW halo. Furthermore, the authors suggest that some satellites have pericenters as close as $\sim 20\,\mathrm{kpc}$, where the strong tidal forces from the central galaxy are important. As we show in Figures 5.1-5.3, depending on when these pericenters take place, the mass of the host was only a fraction of its mass today. Fillingham et al. (2019) use the Phat ELVIS dark matter only simulation suite, which include an analytic disk potential, to statistically sample satellites with similar present-day positions and velocities to make cosmologically informed predictions for what the infall times of the MW's satellites were. The Phat ELVIS simulations account for the growth of the disk potential through abundance matching scaling relations, where each disk has a unique growth rate, and thus, tidally disrupt subhalos that orbit close to the disk (Kelley et al., 2019).

The MW-mass simulations in the FIRE simulation suite do not have an LMC-like analogue at $z = 0$, however, there are four analogues at earlier times, within the last $6 - 7\,\mathrm{Gyr}$ ago, and previous works have used these analogues to study both planar configurations of satellites and

subhalo-stream interactions (Samuel et al., 2021; Barry et al., 2023). Given that the LMC recently fell into the MW's dark matter halo, $\lesssim 2$ Gyr ago (for example Besla et al., 2007; Kallivayalil et al., 2013; Patel et al., 2017), previous orbit modeling studies that include the effects of the LMC to derive orbits of the MW's satellites work well in this recent regime. On the other hand, because we do not include the effects of an LMC in the global potential here in this work, our work informs how well static potential orbit modeling works in the regime before the LMC was a satellite, i.e. $\gtrsim 2$ Gyr ago. As such, both kinds of studies, both with and without an LMC analogue, compliment each other in understanding the full, complex MW formation history.

Acknowledgements

This work relied on the use of NumPy (van der Walt et al., 2011), SciPy (Virtanen et al., 2020), AstroPy (Astropy Collaboration et al., 2022), and Matplotlib (Hunter, 2007). To creatively generate unique color hex codes for some of the figures, we also supplied prompts to the AI drawing model `craiyon`[5] (formerly, DALL·E mini) to create distinct color palettes. Additional color schemes were adopted from the palettes provided by Pantone from the recent JWST press release images.

IBS received support from NASA, through FINESST grant 80NSSC21K1845. AW received support from: NSF via CAREER award AST-2045928 and grant AST-2107772; NASA ATP grant 80NSSC20K0513; HST grants AR-15809, GO-15902, GO-16273 from STScI. We ran simulations using: XSEDE, supported by NSF grant ACI-1548562; Blue Waters, supported by the NSF; Frontera allocations AST21010 and AST20016, supported by the NSF and TACC; Pleiades, via the NASA HEC program through the NAS Division at Ames Research Center.

[5] `https://www.craiyon.com`

CHAPTER 6

Conclusion

Throughout my work in this dissertation, I studied low-mass galaxies at both early cosmic times to understand their role in shaping the formation of galaxies like the MW, and how they leave an imprint on the galaxy that remains today, as well as at late times to understand the dynamics and orbit properties of the satellites that survive the merger history of the MW. To investigate these processes and answer the questions that followed, I used the cosmological, zoom-in simulations of MW-mass galaxies from the FIRE-2 simulations, which reproduce realistic stellar masses, morphologies, and satellite populations. Here, I recapitulate my results from the preceding chapters, and explore an avenue for future analysis.

6.1. Previous Work

In Chapter 2, I explored different ways of quantifying when the main progenitor of the MW may have emerged/formed out of the sea of similar low-mass galaxies at early times. In doing so, I found that the time of formation may be linked through the local environment that galaxies form in, where MW-mass galaxies in pairs, like the MW/M31 pair, typically formed earlier because their dark matter halos formed earlier as previous work had already suggested. I showed that the satellite population today is only a fraction of the progenitor galaxies that formed the MW-mass galaxy and its satellites, thus this work sheds light on future analyses at high redshift focused on the slope of the stellar mass/luminosity function and its evolution across time. This work is also highly relevant to understanding whether or not the progenitor population that built the MW-satellite system is representative of galaxies at high redshift in general, as some studies suggest. Some of the faintest galaxies at high redshift will still not be observable by current and upcoming JWST observations, thus, simulation analyses like this are currently the best options for such predictions.

Chapter 3 focused on a unique population of metal-poor stars that are observed in the MW disk today, on prograde, disk-like orbits. I showed that the overabundance of prograde-to-retrograde metal-poor stars, the prograde bias, is a ubiquitous feature among all but one of the MW-mass hosts in the FIRE-2 simulations. By tracking these stars back in time, I discovered that they primarily formed in low-mass galaxies that were not the proto-MW, and although a handful of mergers deposited them over time, a single gas-rich LMC/SMC-mass merger is responsible for a significant fraction. This merger deposited the metal-poor stars and a significant amount of gas on an orbit that sometimes contributed to the formation and stabilization of the disk itself. This work was one of the first predictions for this feature in the MW disk, and agrees well with other results from another simulation group, strengthening the idea that this is common amongst MW-mass galaxies, and overall, another piece of evidence for the hierarchical formation paradigm.

I then shifted my focus in Chapter 4 to study the present-day dynamics and orbit histories of the satellite populations of MW-mass galaxies. My results qualitatively agree with the idea that some satellite galaxies are 'pre-processed' before they become satellites of their more massive host galaxies, as other previous studies have suggested, and is seen today in the LMC which hosts its own satellites as it orbits throughout the MW. I also showed that one cannot accurately infer a satellite galaxy's time of infall into the MW-mass host from its present-day properties, such as its distance from the host or its current stellar mass, given the wide range in possible infall times for a particular distance/mass. One common assumption is that as a satellite galaxy orbits within the dark matter halo of its MW-mass host, it experiences dynamical friction and its orbit must always shrink over time, however, as I showed, for the satellites that have orbited multiple times, it is more common that their most recent pericenter is not the smallest. The smallest pericenter a satellite experiences often happens much earlier than its most recent, and this increase in pericenter distance is a result of either a gradual or rapid growth in its specific orbital angular momentum. This change in angular momentum may be attributed to the gradual growth of the MW-mass host itself, or through dynamical interactions with other satellite galaxies in the outer dark matter halo, or some

combination of processes, although more work is needed to definitively address this. Thus, my work challenges some commonly held assumptions in the orbit modeling community.

Finally, in Chapter 5, I focused on better understanding and quantifying the utility of simple orbit modeling techniques, and how well they work in deriving satellite galaxy orbit histories. I showed that orbital 'invariants' like the total energy and specific angular momentum of a satellite galaxy are not conserved properties across time, and rather, that the uncertainty in each is non-negligible only a few Gyr back. The axisymmetric model recovers properties that occurred in the recent past, such as the timing, distance, and velocity of the recent pericenter, better than properties that occurred farther back in time, like the minimum pericenter distance and lookback time of satellite infall. Even when the median trends between the model and simulation agree well, the typical uncertainties in orbit properties can be a factor of two or more, and given the variation across the entire sample, one should not approximate a property for a single galaxy from the median trend. Because we model the mass/potential of the MW-mass host to within a few percent, these results represent a lower-limit on the accuracy of static potential orbit modeling, and accounting for uncertainties in mass will only increase the unreliability.

6.2. Future Work

Many recent studies use orbit modeling to infer the orbit histories of satellites within the LG, however, they often implement static potentials or compare results against dark matter only simulations (for example Besla et al., 2007; Kallivayalil et al., 2013; Patel et al., 2017; Fritz et al., 2018a; Fillingham et al., 2019; Patel et al., 2020). The major limitations in these studies include not accounting for a time-dependent potential, neglecting the gravitational influence of the central galaxy, or that they do not account for the effects from large scale structure in a cosmological context.

The FIRE suite of simulations are excellent tools for this kind of analysis because (1) they are cosmological and thus, account for the effects from large scale structure, (2) model hydrodynamic effects including star formation, stellar feedback, and by consequence the gravitational contribution

from the central galaxy, and (3) account for realistic growth histories of MW-mass galaxies and their satellite populations. These simulations therefore have the advantage over dark matter only simulations because of the inclusion of baryonic processes, and have an advantage over purely analytic orbit modeling techniques because they model the true time-dependent mass/potential growth of the MW-mass host. Furthermore, based on the work I presented in Chapter 5, the total orbital energy and angular momentum of satellites are not conserved across time, as static orbit modeling techniques might suggest. With the FIRE suite, I can match satellites in the simulations to the actual satellites of the MW or M31 through their stellar mass, distance, radial velocity, or proper motion. By doing so, I can statistically sample the orbit histories of each satellite to measure cosmologically informed probabilities of their pericenter properties or infall times.

This is not to say that there are no limitations with the FIRE simulations. Throughout my work I used the 14 MW-mass hosts, with a combined satellite population of $\approx$ 500 satellites, spanning stellar masses from $M_{\rm star} \sim 3 \times 10^4 - 10^9\,{\rm M}_\odot$. This sample size of satellites may not be sufficient to statistically sample each satellite around the MW in their stellar mass or 6D phase-space properties. The mass resolution of star particles in the Latte and ELVIS on FIRE suites is $m_{\rm star} = 3500 - 7100\,{\rm M}_\odot$, and given that we can minimally resolve these low-mass galaxies with ~ 6 star particles, I will also not be able to model the orbits of galaxies below $M_{\rm star} < 10^4\,{\rm M}_\odot$. Finally, as previous research suggests, the contribution to the global galactic potential from the LMC is non-negligible for some satellites, and the FIRE simulations do not have perfect LMC analogs at $z = 0$.

Because of the limitations in sample size and stellar mass, I may use dark matter only simulations which will both increase the number of satellites and allow me to use a stellar mass to halo mass relation to sample lower mass satellites. However, without the additional gravitational potential from the disk, I run the risk of including satellites that otherwise would have been completely tidally disrupted when orbiting too close to the central galaxy. One workaround to this is using the Phat ELVIS simulations (Kelley et al., 2019), or generating similar simulations, which include an

analytic disk-like potential in dark matter only hosts. The lack of stellar feedback within the satellite also induces a 'survivor-ship' bias, given that feedback ultimately reduces the dark matter density of a subhalo by driving fluctuations to the potential, making it more susceptible to disruption (El-Badry et al., 2016, 2017; Sales et al., 2022). However, this is a second-order effect to the destruction of subhalos as previous work suggests (Garrison-Kimmel et al., 2017b).

By leveraging current HST and Gaia data along with the cosmological simulations from FIRE, I will generate cosmologically informed posteriors on the orbit properties of the satellites within the LG. This method is advantageous over current orbit modeling techniques given the limitations I presented in Chapter 5, and the assumptions that come with static potentials. Because there is more recent work in the literature on the satellites of the MW, I will first apply this framework to all satellites of M31, and then to the satellites of the MW and compare my results with previous studies. Thus, my work both builds upon the previous orbit modeling studies and makes predictions for future studies focused on the orbital dynamics of M31's satellite population. I finally plan to make any new simulations, merger trees, or python pipelines publicly available as a resource for others in the scientific community.

APPENDIX A

Chapter 4 Appendices

A.1. Trends with Peak Halo Mass

In Section 4.4, we investigated trends of satellite orbital dynamics and histories as a function of satellite $M_{\rm star}$, which is the mass most directly observable. However, from Figure 4.1, the DM (sub)halo mass of a satellite is $10^2 - 10^4\times$ larger, so it is the one most important for dynamics. Here we investigate the same trends but as a function of a satellite's peak halo mass, $M_{\rm halo,peak}$.

We select *all* subhalos with $M_{\rm halo,peak} > 10^8\,{\rm M}_\odot$, which includes both luminous and dark subhalos (with no stars). Thus, given the extrapolated abundance matching relations of Moster et al. (2013); Garrison-Kimmel et al. (2017a); Behroozi et al. (2020) in comparison to our stellar mass selected sample in Figure 4.1, this includes lower-mass subhalos that likely would host ultra-faint galaxies whose stellar masses our baryonic simulations do not natively resolve. For reference, the fraction of satellites in each mass bin that are luminous in our simulations is: 1 per cent for $M_{\rm halo,peak} = 10^{8-8.5}\,{\rm M}_\odot$, 14 per cent for $M_{\rm halo,peak} = 10^{8.5-9}\,{\rm M}_\odot$, 60 per cent for $M_{\rm halo,peak} = 10^{9-9.5}\,{\rm M}_\odot$, 92 per cent for $M_{\rm halo,peak} = 10^{9.5-10}\,{\rm M}_\odot$, and 100 per cent above $M_{\rm halo,peak} > 10^{10}\,{\rm M}_\odot$. Compared to our stellar mass selection, this increases our satellite sample size by a factor of ≈ 8.5.

Figure A.1 caption: Similar to Figure 4.2, the orbital dynamics of satellite galaxies at $z = 0$, here versus their peak halo mass, $M_{\rm halo,peak}$, for all satellites (luminous and dark). **Top**: Median orbital total velocity is nearly constant with $M_{\rm halo,peak}$ at $\approx 135\,{\rm km\,s^{-1}}$ and 68th percentile ranging from $71 - 188\,{\rm km\,s^{-1}}$. **Middle**: Median specific orbital total energy, E, is nearly constant at $-0.9\times10^4\,{\rm km^2\,s^{-2}}$ with a 68th percentile range of -2.6 to $-0.2\times10^4\,{\rm km^2\,s^{-2}}$. Notably, the dependence on $M_{\rm halo,peak}$ is event flatter than with $M_{\rm star}$ in Figure 4.2, though these quantities

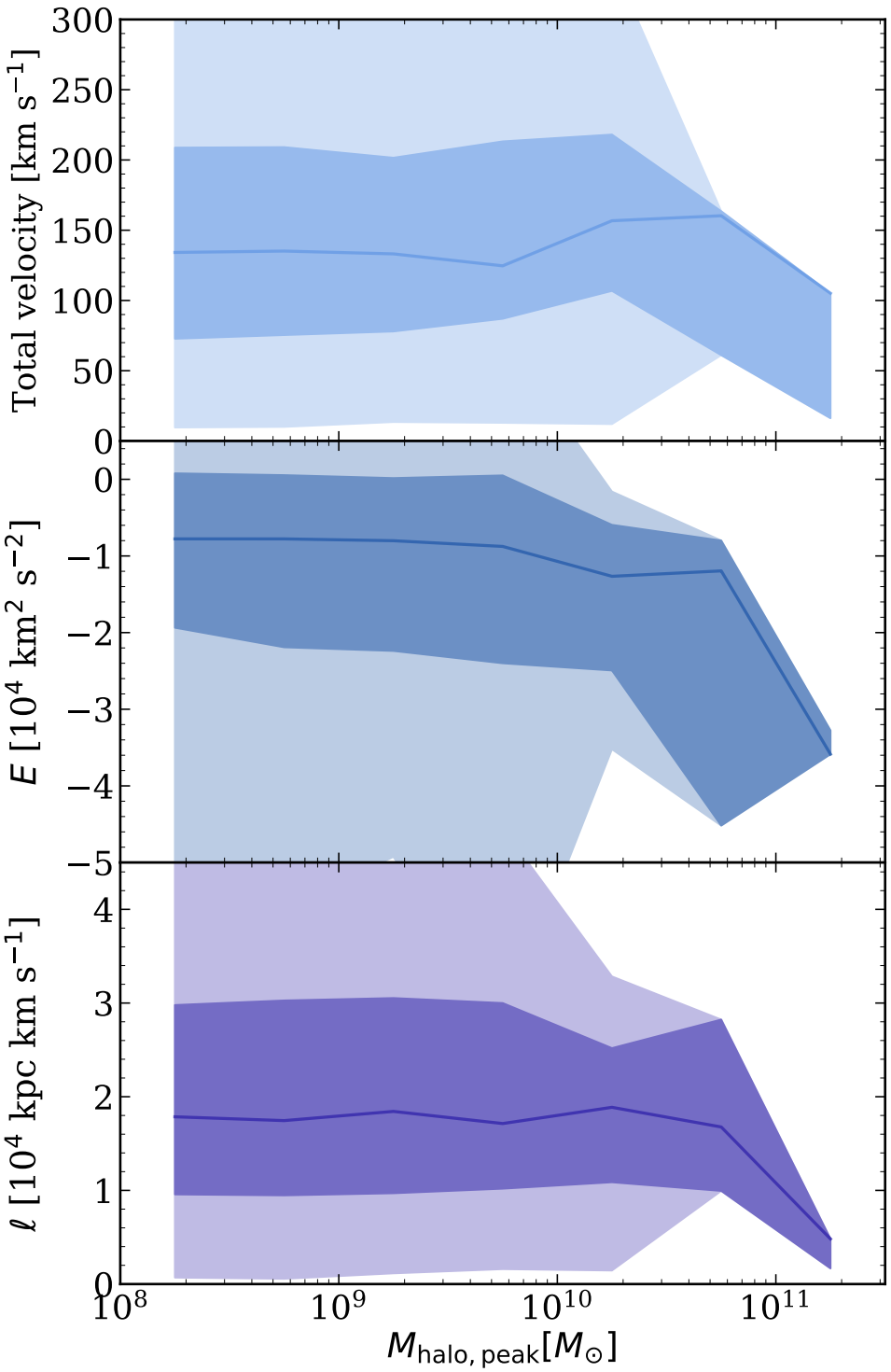

FIGURE A.1. See figure caption in the text.

all decline rapidly at $M_{\rm halo,peak} \gtrsim 5 \times 10^{10}\,{\rm M}_\odot$, likely because of the increasing importance of dynamical friction for sufficiently massive satellites. **Bottom**: Median orbital specific angular momentum, ℓ, is nearly constant at $\ell \approx 1.6 \times 10^4\,{\rm kpc\,km\,s^{-1}}$ with a 68th percentile range of $0.9 - 2.6 \times 10^4\,{\rm kpc\,km\,s^{-1}}$.

Because (sub)halo mass is the most relevant dynamically, but a satellite with a given $M_{\rm halo,peak}$ hosts a range of stellar masses given the scatter in the SMHM relation, trends with $M_{\rm star}$ tend to be noisier. Figures A.1, A.2, and A.3 all show qualitatively similar trends to those in Section 4.4. In particular, the trends at low halo mass in Figure A.1 show relatively flat dependence, and at $M_{\rm halo,peak} \gtrsim 10^{10.5}\,{\rm M}_\odot$, we see a more pronounced decline given the stronger dynamical friction at these masses. Trends with the lookback times of both infall metrics and the pericentre lookback times all qualitatively show similar results and offsets in Figure A.2, and the number of pericentric passages agrees with the stellar mass selection, though it is shifted to slightly smaller $N_{\rm peri} \approx 2$ for the smallest subhalos, compared to $N_{\rm peri} \approx 2.5$ at our lowest stellar masses. We do not show trends of pericentre distance, given the lack of a strong dependence on $M_{\rm halo,peak}$, but we compare $d_{\rm peri,min}$ for satellites in baryonic versus DMO simulations in Appendix A.2. In summary, the trends using this halo-mass selected sample are qualitatively similar to the results presented throughout Section 4.4. Furthermore, our results here imply similar trends for ultra-faint galaxies, where no halo capable of hosting an ultra-faint galaxy, $M_{\rm halo,peak} \approx 10^8\,{\rm M}_\odot$, was a satellite of the MW-mass host halo progenitor during the epoch of reionization, $z \gtrsim 6$. Similar to the results in our stellar-mass selected sample, we also find that < 1 per cent of the satellites in this halo-selected sample were members of a more massive halo during reionization.

Figure A.2 caption: Similar to Figure 4.4, now as a function of peak halo mass, $M_{\rm halo,peak}$, for all satellites (both luminous and dark) at $z = 0$, showing stronger and smoother trend with $M_{\rm halo,peak}$. **Top**: Similar with the trends with $M_{\rm star}$, lower-mass satellites fell into any more massive halo, or their MW-mass halo, earlier than higher-mass satellites, and satellites with $M_{\rm halo,peak} \lesssim 10^{10}\,{\rm M}_\odot$ typically fell into their MW-mass halo $\sim 1 - 2\,{\rm Gyr}$ after falling into any more massive halo. As the

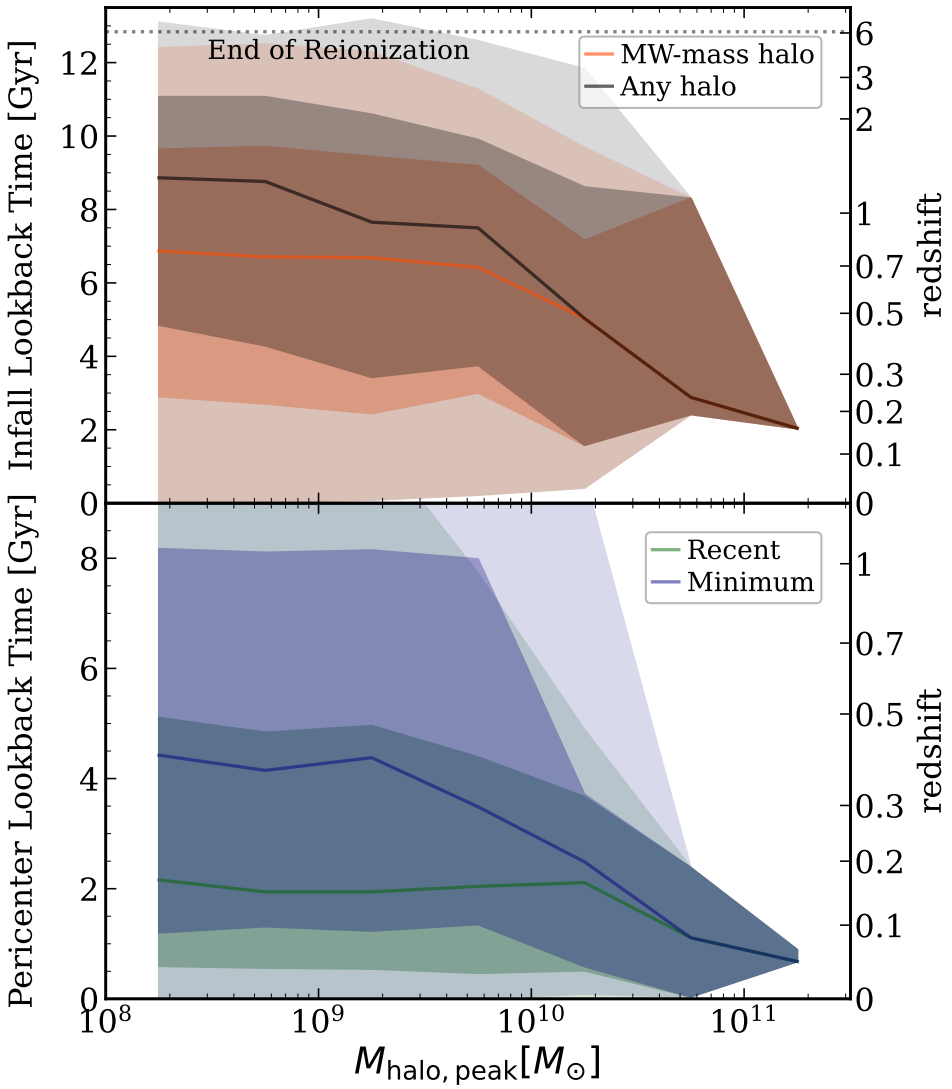

FIGURE A.2. See the figure caption in the text.

full distribution shows, no halos in our sample were satellites of their MW-mass hosts during the epoch of reionization ($z \gtrsim 6$), even down to the ultra-faint regime, $M_{\mathrm{halo,peak}} \approx 10^8\,\mathrm{M}_\odot$, and less than 1 per cent of satellites were members of another more massive halo during that time. **Bottom**: As in Figure 4.4, the lookback times of both the minimum and most recent pericentres occur earlier for satellites at lower $M_{\mathrm{halo,peak}}$. The median $t^{\mathrm{lb}}_{\mathrm{peri,min}}$ decreases from ≈ 4.5 Gyr ago for lower-mass

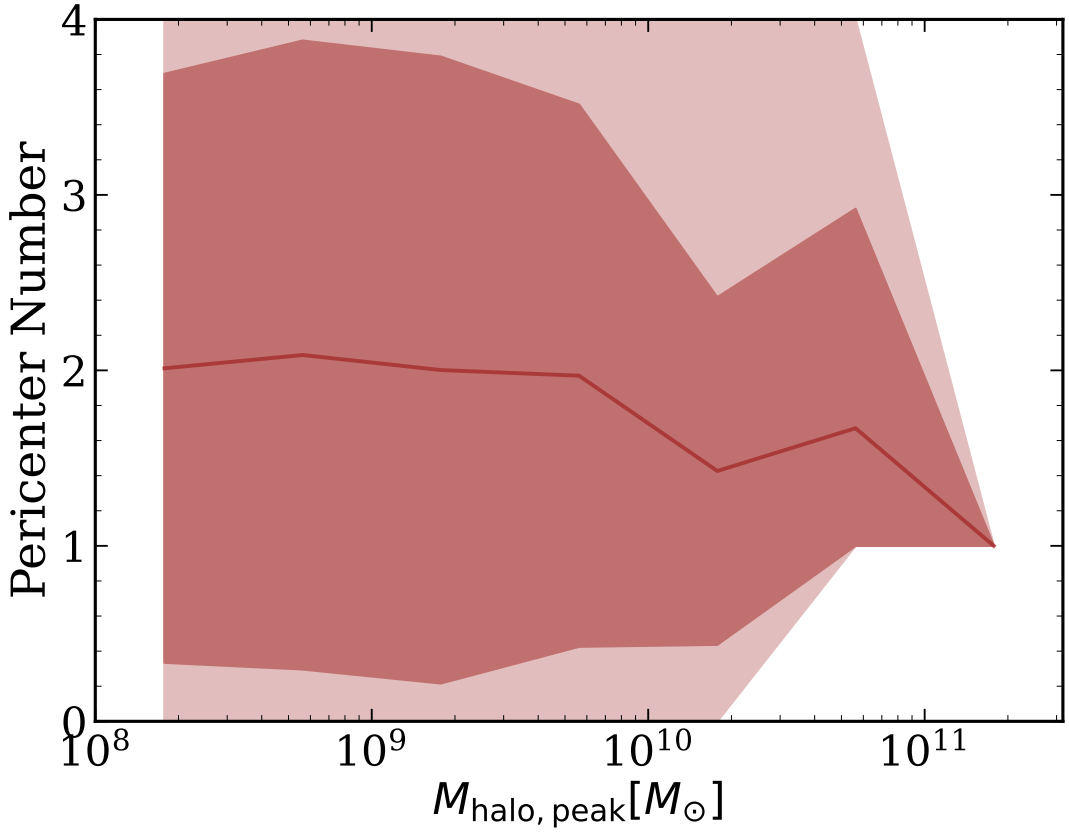

FIGURE A.3. Similar Figure 4.6 (top right), now showing the number of pericentric passages about the MW-mass host versus peak halo mass, $M_{\rm halo,peak}$. Now the trend with mass slightly weaker, decreasing from 2 to 1 with increasing $M_{\rm halo,peak}$, because lower-mass satellites fell into the MW-mass halo earlier and on smaller orbits and do not experience as strong of dynamical friction. The full distribution (clipped for visual clarity) is $0-12$ pericentres at $M_{\rm halo,peak} \approx 10^{8.5}\,{\rm M}_\odot$.

satellites to ≈ 0.75 Gyr ago at our highest masses. By contrast, the median $t^{\rm lb}_{\rm peri,rec}$ is much flatter with $M_{\rm halo,peak}$.

A.2. Baryonic versus dark-matter-only simulations

Here we compare our results from our FIRE-2 baryonic simulations against satellites in dark matter-only (DMO) simulations of the same halos, to understand the effects of baryons and contextualize previous results based on DMO simulations, given that many previous works investigated

satellite orbits and infall histories in DMO simulations (for example Wetzel et al., 2015; Bakels et al., 2021; D'Souza & Bell, 2021; Robles & Bullock, 2021; Ogiya et al., 2021), which, among other things, do not model the potential from a central galaxy. Furthermore, stellar feedback in more massive satellites can reduce their inner dark-matter densities, making them more susceptible to tidal disruption (for example Bullock & Boylan-Kolchin, 2017). With no tidal forces from the central galaxy and with less dense dark-matter cusps within the subhalos, tidal disruption can be much stronger in baryonic simulations. Recent studies also have used DMO simulations with an embedded disk-like potential (for example Kelley et al., 2019; Rodriguez Wimberly et al., 2019; Fillingham et al., 2019; Robles & Bullock, 2021).

We compare only simulations that have DMO counterparts at all snapshots, which comprises the 7 MW-mass hosts in isolated environments (names beginning with 'm12') in Table 4.1. As in Appendix A.1, for all simulations we select all satellites with $M_{\rm halo,peak} > 10^8\,{\rm M}_\odot$, which includes both luminous and dark satellites in the baryonic simulations. In the DMO simulations, we re-normalize $M_{\rm halo,peak}$ to account for the loss of baryons by multiplying by $1 - f_{\rm b}$, where $f_{\rm b} = \Omega_{\rm baryon}/\Omega_{\rm matter}$ is the cosmic baryon fraction. The total number of satellites in the DMO simulations is $\approx 1.6\times$ higher.

Figure A.4 shows the lookback times of 'first' infall into any other more massive halo, $t^{\rm lb}_{\rm infall,any}$ (top left), specific angular momentum, ℓ (top right), the smallest pericentre experienced, $d_{\rm peri,min}$ (bottom left), and the number of pericentric passages about the MW-mass host, $N_{\rm peri}$ (bottom right), for satellites in the baryonic (red) and DMO (black) simulations. Solid lines show the median and the dark and light shaded regions show the 68th percentile and full distribution, respectively.

Satellites in the DMO simulations do not feel the gravity of a central galaxy, so they experience weaker tidal stripping and disruption, even if they fell in early or orbit closer to the center of the halo. Thus, the (surviving) satellites in the DMO simulations generally fell in $0.5 - 3\,{\rm Gyr}$ earlier than in the baryonic simulations. As a result of the surviving population falling in earlier, satellites in DMO simulations also orbit at smaller distances; they were able to orbit closer to the center of the

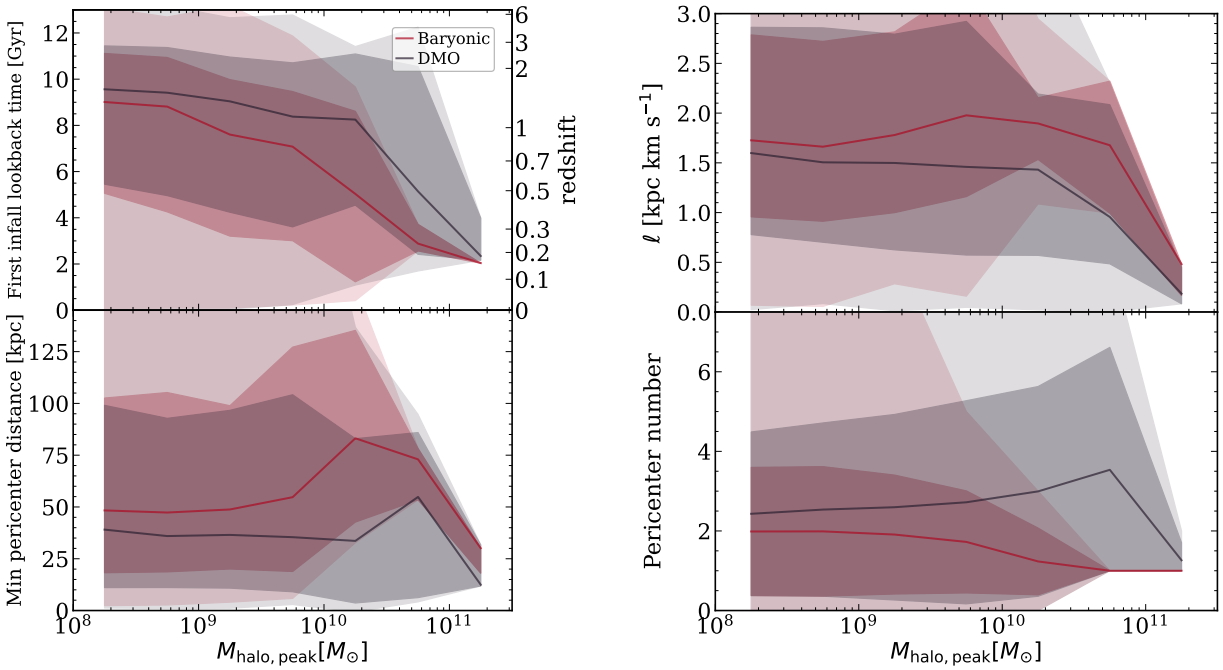

FIGURE A.4. Comparing satellite orbital properties versus $M_{\rm halo,peak}$ in the baryonic simulations (red) against dark matter-only (DMO) simulations (black) of the same systems, for satellites at $z = 0$ within the isolated MW-mass halos. **Top left**: Satellites in the DMO simulations fell into a more massive halo $0.5-3.5$ Gyr earlier than in the baryonic simulations. **Bottom left**: The median minimum pericentre distance, $d_{\rm peri,min}$, is smaller in DMO simulations, ranging typically from $10-55$ kpc as compared with $30-85$ kpc in the baryonic simulations. **Top right**: The orbital specific angular momentum, ℓ, is smaller in DMO simulations than in baryonic simulations. **Bottom right**: The mean number of pericentric passages about the MW-mass host is smaller in the baryonic simulations than in DMO. Also, $N_{\rm peri}$ increases slightly with satellite mass in the DMO simulations, while it decreases monotonically in the baryonic simulations. The primary reason for these differences is that the central galaxy in the baryonic simulations induces stronger tidal stripping and disruption on satellites that orbit near it, leading to a *surviving* population that fell in more recently, experienced fewer and larger-distance pericentres, and has more orbital angular momentum.

MW-mass halo without becoming tidally disrupted, as the bottom left panel shows. Furthermore, surviving satellites have lower ℓ in DMO simulations, given that satellites with smaller ℓ in the baryonic simulations are likely to be tidally disrupted (for example Garrison-Kimmel et al., 2017b). Finally, because satellites in DMO simulations fell in earlier and orbit at smaller distances, they

completed more pericentric passages (bottom right panel). We also see a small increase in $N_{\rm peri}$ with $M_{\rm halo,peak}$, likely because higher-mass satellites in DMO simulations in particular can survive longer than in the presence of a central galaxy.

Our results agree with Garrison-Kimmel et al. (2017b), who compared subhalo populations between DMO and FIRE-2 baryonic simulations using 2 of the same systems that we analyze (m12i and m12f). They also tested the results of using a DMO simulation with an analytic galaxy potential embedded within the host halo, finding good agreement with the baryonic simulations, which implies that the most important effect in the baryonic simulations is additional gravitational effect of the MW-mass galaxy. They showed that the number of subhalos between the different types simulations converges for subhalos that orbit farther away from the center of the MW-mass halo. Thus, differences between DMO and baryonic simulations are largest for subhalos that orbit closer to the center of the host, where these subhalos get preferentially disrupted in the baryonic simulations, and result in a satellite population with a larger fraction on more tangential orbits, with higher specific angular momentum.

APPENDIX B

Chapter 5 Appendix

B.1. Modeling the Mass/Potential Profiles

We integrate the satellite galaxy orbits using the galactic dynamics Python package `galpy` (Bovy, 2015), which allows users to define custom MW-mass host potentials. Thus, we first model the enclosed mass profiles of the MW-mass hosts in the simulations as a sum of the disk and halo components, to derive parameters for their global potentials in `galpy`.

B.1.1. Modeling the disk. We model the disk as the sum of two double exponential disks, one for the inner disk and bulge contribution, and one for the outer disk region. The full density profile of the disk is:

$$\rho(R, Z) = A_{\rm disk}^{\rm inner}\ e^{-R/r_{\rm disk}^{\rm inner} - |Z|/h_z}\ + A_{\rm disk}^{\rm outer}\ e^{-R/r_{\rm disk}^{\rm outer} - |Z|/h_z} \tag{B.1}$$

where $A_{\rm disk}^{\rm inner}$ & $A_{\rm disk}^{\rm outer}$ are amplitudes with units of mass density ($M_\odot/\,{\rm kpc}^3$), $r_{\rm disk}^{\rm inner}$ & $r_{\rm disk}^{\rm outer}$ are disk scale lengths (kpc), and h_z is the disk scale height (kpc). Given that satellite orbits are more sensitive to the enclosed *mass*, rather than the local density field, we first integrate out the vertical component of Equation B.1, then integrate over cylindrical R to obtain the disk enclosed mass given by

$$M(<R) = 4\pi A h_z r_{\rm inner}\left[r_{\rm inner} - e^{-R/r_{\rm inner}}(r_{\rm inner} + R)\right], \tag{B.2}$$

and a similar term for the outer disk component. By converting the density profiles to enclosed mass profiles, we ensure that the total mass is conserved when integrated over all space. We fit and model the disk with star particles and cold gas ($T < 10^5 K$) within $R = [0, 20]$ kpc and $|Z| < 3$ kpc. Including mass within $|Z| < 4, 5$ did not significantly change the derived parameters,

and artificially boosting the disk amplitudes to account for additional baryonic mass beyond this geometric selection resulted in enclosed masses that were too large.

We also checked how fitting the disk to a Miyamoto-Nagai or single exponential profile compared to the simulations, however, we obtained better fits to the enclosed mass with the method outlined above. Furthermore, the orbits of satellites at distances well beyond the disk are not highly sensitive to the precise distribution of mass enclosed.

B.1.2. Modeling the dark matter halo. To provide more flexibility to the mass profile, we use a generalized form of the spherical Navarro-Frenk-White (NFW, Navarro et al., 1996) halo density profile, defined as:

$$\rho(r) = \frac{A_{\rm halo}}{4\pi a_{\rm halo}^3} \frac{1}{(r/a_{\rm halo})^{\alpha}(1 + r/a_{\rm halo})^{\beta-\alpha}} \tag{B.3}$$

where $A_{\rm halo}$ is an amplitude (given in $\rm M_\odot$), $a_{\rm halo}$ is the halo scale radius (given in kpc), and α and β are slopes to the inner and outer regions of the density profile. We then integrate the analytic form of $\rho(r)$ to convert this to an enclosed mass profile, given by:

$$M(< r) = \frac{A_{\rm halo}}{3-\alpha}\left(\frac{r}{a}\right)^{3-\alpha} {}_2F_1(3-\alpha, -\alpha+\beta; 4-\alpha; -r/a_{\rm halo}) \tag{B.4}$$

where ${}_2F_1$ is the Gauss hypergeometric function. We model and fit this profile with dark matter and hot gas ($T > 10^5 K$) particles within $r < 10$ kpc, and *all* particles (dark matter, hot+cold gas, and stars) from 10 kpc $< r <$ 500 kpc. We obtained similar derived parameters by selecting particles within $r < 300, 350, 400$ kpc, however, because many of the satellite galaxies are splashback satellites (see Santistevan et al., 2023), we choose to keep the fits out to 500 kpc. We additionally fit the halos to the regular NFW profile, where $\alpha = 1$ and $\beta = 3$, however, given the extra complexity of the two-power NFW model, we obtained better agreement between the data and the derived model.

B.1.3. Full enclosed mass profile. The parameters we derived from fitting the analytic density/mass profiles above are used in their respective potentials in `galpy`. In `galpy`, we specifically use the spherically symmetric 'TwoPowerSphericalPotential' for the DM halo, and axisymmetric 'DoubleExponentialDiskPotential' for the inner and outer disks, and input all three to define the custom global potential for each of our simulated host galaxies to integrate the orbits of their satellite populations. Table B.1 lists the complete set of derived parameters.

TABLE B.1. Best-fit parameters to the double-exponential disk profile (Equations B.1-B.2) and the generalized NFW profile for the halo (Equations B.3-B.4). Columns: Name of host; halo amplitude, $A_{\rm halo}$; halo scale radius, $a_{\rm halo}$; halo inner slope, α; halo outer slope, β; disk inner amplitude, $A_{\rm disk}^{\rm inner}$; disk inner scale length, $r_{\rm inner}$; disk outer amplitude, $A_{\rm disk}^{\rm outer}$; disk outer scale length, $r_{\rm outer}$; disk scale height, h_z.

Name	$A_{\rm halo}$ [10^{11} M$_\odot$]	$a_{\rm halo}$ [kpc]	α	β	$A_{\rm disk}^{\rm inner}$ [10^{9} M$_\odot$ kpc^{-3}]	$r_{\rm disk}^{\rm inner}$ [kpc]	$A_{\rm disk}^{\rm outer}$ [10^{8} M$_\odot$ kpc^{-3}]	$r_{\rm disk}^{\rm outer}$ [kpc]	$h_{\rm z}$ [kpc]
m12m	3.78	17.32	1.57	2.78	6.47	0.79	7.98	4.41	0.64
Romulus	5.30	6.65	0.00	2.87	8.31	0.86	2.10	8.04	0.55
m12b	3.84	16.89	1.48	2.82	18.0	0.65	6.08	4.24	0.51
m12f	3.65	14.77	1.45	2.74	8.39	0.84	2.11	7.40	0.54
Thelma	4.49	8.84	0.40	2.90	3.30	1.04	1.98	6.28	0.75
Romeo	8.99	29.91	1.47	3.23	4.61	1.02	2.25	6.64	0.55
m12i	5.70	32.15	1.58	2.99	9.10	0.78	3.87	4.56	0.55
m12c	7.93	24.37	0.90	3.05	7.05	0.71	6.41	3.65	0.57
m12w	2.99	22.87	1.61	2.73	5.53	0.68	11.7	2.19	0.67
Remus	8.32	36.49	1.56	3.21	4.68	0.90	1.60	6.50	0.54
Juliet	2.85	9.57	0.87	2.82	6.97	0.77	1.16	6.32	0.55
Louise	2.30	6.41	0.29	2.76	1.92	0.99	0.54	9.36	0.58
m12z	3.22	33.96	1.58	2.69	0.59	0.40	1.73	2.94	1.54

We show how well we fit each of these enclosed mass profiles in Figure B.1, where we show the ratio of the halo+disk analytic model using the parameters in Table B.1 to the enclosed masses in the simulations; a 1:1 ratio is presented as the dotted horizontal line. We plot the median enclosed mass profile ratio across all 13 galaxies in the solid black line, and we also show the 68th and 95th percentile ranges in the dark and light lavender shaded regions, respectively. The median trend shows that for all distances we present, the enclosed mass is modeled to within 3-4 per cent or so, and the 68th percentile range is within 5 per cent beyond 10 kpc. Within 10 kpc, the 68th and 95th percentiles of the enclosed mass ratios span both higher and lower than what we present in the y-axis, however, very few satellites ever orbit at such small distances. Only $\sim$ 5 per cent of satellites orbit within 10 kpc, thus, obtaining more precise enclosed mass profiles of these inner regions of the galaxy are not as crucial as the outer regions.

Individually, the model of the two-component disk is modeled to within $\sim$ 4 per cent between $r = 2 - 20$ kpc and the DM halo is modeled to within $\approx 10 - 20$ per cent. One galaxy, m12c, experienced a significant merger ($\sim 3:1$ mass ratio) around ~ 8.97 Gyr ago, thus modeling this host with symmetric analytic models does not capture the complexity in its mass distribution. Thus, the total enclosed mass in m12c is fit to within $\sim$ 15 per cent.

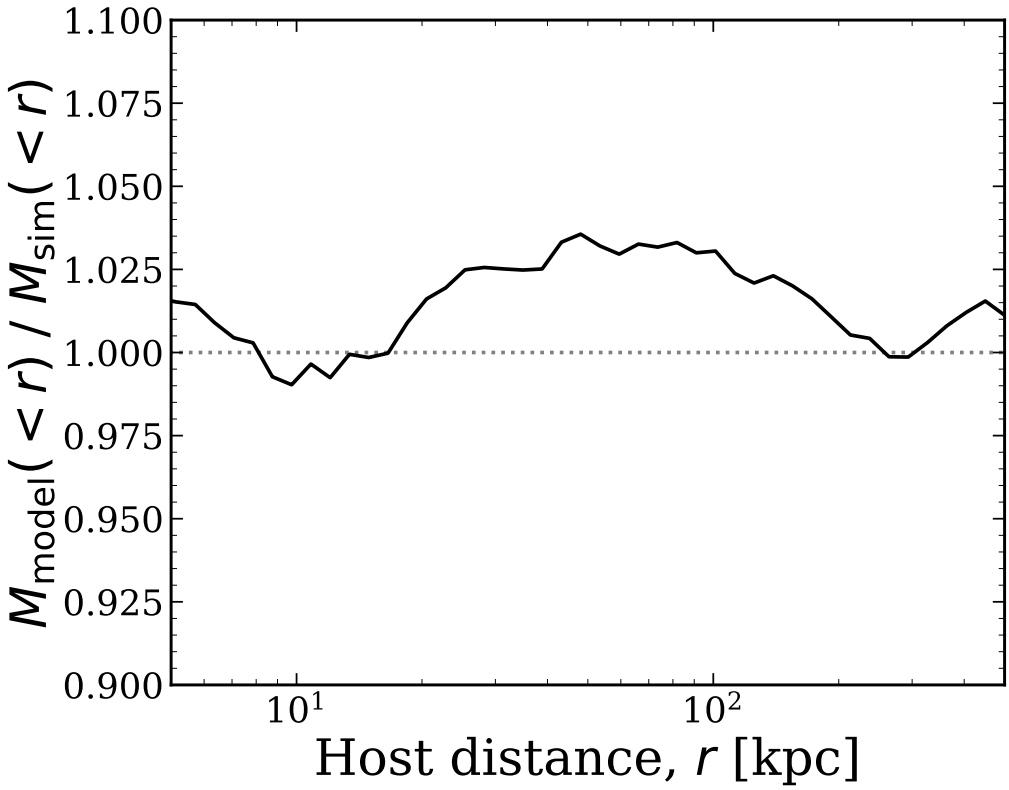

FIGURE B.1. Ratio of the enclosed total mass in the best-fit model to that in the simulation, as a function of distance from the center of the MW-mass host, r, at $z = 0$. We fit a double-exponential profile for the inner and outer disk (Equation B.1-B.2) and a generalized NFW profile for the halo (Equations B.3-B.4). Table B.1 lists all fit parameters. Black line shows the median and the dark and light shaded regions show the 68th and 95th percentile scatter across our 13 hosts. The median agrees to within < 4 per cent at all radii and agrees to $\lesssim 1$ per cent for the 'total' halo mass within the virial radius ($r = 344 - 472$ kpc). The ratio shows larger host-to-host scatter at $\lesssim 10$ kpc, but only ~ 5 per cent of satellites in our sample orbit this close. This level of agreement ensures that modeling the total (spherically symmetric) host mass profile at $z = 0$ is not a significant source of error for our orbit modeling.

Bibliography

Adhikari S., Dalal N., Chamberlain R. T., 2014, JCAP, 2014, 019

Amorisco N. C., 2017, MNRAS, 464, 2882

Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698

Arora A., Sanderson R. E., Panithanpaisal N., Cunningham E. C., Wetzel A., Garavito-Camargo N., 2022, ApJ, 939, 2

Astropy Collaboration et al., 2022, ApJ, 935, 167

Athanassoula E., Rodionov S. A., Peschken N., Lambert J. C., 2016, ApJ, 821, 90

Bakels L., Ludlow A. D., Power C., 2021, MNRAS, 501, 5948

Barbá R. H., Minniti D., Geisler D., Alonso-García J., Hempel M., Monachesi A., Arias J. I., Gómez F. A., 2019, ApJL, 870, L24

Barbuy B., Chiappini C., Gerhard O., 2018, AR&AA, 56, 223

Barnes J. E., 2004, MNRAS, 350, 798

Barnes J. E., Hernquist L., 1992, AR&AA, 30, 705

Barnes J., Hut P., 1986, Nature, 324, 446

Barry M., Wetzel A., Chapman S., Samuel J., Sanderson R., Arora A., 2023, arXiv e-prints, p. arXiv:2303.05527

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109

Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Behroozi P., et al., 2020, MNRAS, 499, 5702

Bellardini M. A., Wetzel A., Loebman S. R., Faucher-Giguère C.-A., Ma X., Feldmann R., 2021, MNRAS, 505, 4586

Bellardini M. A., Wetzel A., Loebman S. R., Bailin J., 2022, MNRAS, 514, 4270

Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, Monthly Notices of the Royal Astronomical Society, 478, 611

Bennet P., Alfaro-Cuello M., Pino A. d., Watkins L. L., van der Marel R. P., Sohn S. T., 2022, ApJ, 935, 149

Bernard E. J., Schultheis M., Di Matteo P., Hill V., Haywood M., Calamida A., 2018, MNRAS, 477, 3507

Bertschinger E., 1985, ApJS, 58, 39

Besla G., Kallivayalil N., Hernquist L., Robertson B., Cox T. J., van der Marel R. P., Alcock C., 2007, ApJ, 668, 949

Bird J. C., Kazantzidis S., Weinberg D. H., Guedes J., Callegari S., Mayer L., Madau P., 2013, ApJ, 773, 43

Bird J. C., Loebman S. R., Weinberg D. H., Brooks A. M., Quinn T. R., Christensen C. R., 2021, MNRAS, 503, 1815

Bland-Hawthorn J., Gerhard O., 2016, AR&AA, 54, 529

Blanton M. R., et al., 2017, AJ, 154, 28

Bonaca A., Geha M., Küpper A. H. W., Diemand J., Johnston K. V., Hogg D. W., 2014, ApJ, 795, 94

Bonaca A., Conroy C., Wetzel A., Hopkins P. F., Kereš D., 2017, ApJ, 845, 101

Bonaca A., et al., 2021, ApJL, 909, L26

Bournaud F., Combes F., Jog C. J., 2004, A&A, 418, L27

Bournaud F., Jog C. J., Combes F., 2005, A&A, 437, 69

Bouwens R. J., et al., 2015, ApJ, 803, 34

Bovy J., 2015, ApJS, 216, 29

Bovy J., Rix H.-W., 2013, ApJ, 779, 115

Bovy J., Rix H.-W., Schlafly E. F., Nidever D. L., Holtzman J. A., Shetrone M., Beers T. C., 2016a, ApJ, 823, 30

Bovy J., Bahmanyar A., Fritz T. K., Kallivayalil N., 2016b, ApJ, 833, 31

Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150

Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., 2014, MNRAS, 443, L44

Boylan-Kolchin M., Weisz D. R., Johnson B. D., Bullock J. S., Conroy C., Fitts A., 2015, MNRAS, 453, 1503

Brook C. B., Kawata D., Scannapieco E., Martel H., Gibson B. K., 2007, ApJ, 661, 10

Brooks A. M., Zolotov A., 2014, ApJ, 786, 87

Bryan G. L., et al., 2014, ApJS, 211, 19

Buck T., Obreja A., Macciò A. V., Minchev I., Dutton A. A., Ostriker J. P., 2019, arXiv e-prints, p. arXiv:1909.05864

Buck T., Obreja A., Macciò A. V., Minchev I., Dutton A. A., Ostriker J. P., 2020, MNRAS, 491, 3461

Bullock J. S., Boylan-Kolchin M., 2017, AR&AA, 55, 343

Bullock J. S., Johnston K. V., 2005, The Astrophysical Journal, 635, 931

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2001, The Astrophysical Journal, 548, 33

Carlesi E., Hoffman Y., Gottlöber S., Libeskind N. I., Knebe A., Yepes G., Pilipenko S. V., 2020, MNRAS, 491, 1531

Carney B. W., Laird J. B., Latham D. W., Aguilar L. A., 1996, AJ, 112, 668

Carter C., et al., 2021, ApJ, 908, 208

Chandrasekhar S., 1943, ApJ, 97, 255

Chaplin W. J., et al., 2014, ApJS, 210, 1

Chiaki G., Wise J. H., 2019, MNRAS, 482, 3933

Chiba M., Beers T. C., 2000, AJ, 119, 2843

Ciardi B., Ferrara A., 2005, SSR, 116, 625

Colberg J. M., et al., 2000, MNRAS, 319, 209

Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720

Conroy C., et al., 2019, ApJ, 883, 107

Cooper A. P., et al., 2010, MNRAS, 406, 744

Cordoni G., et al., 2020, MNRAS,

Corlies L., Johnston K. V., Wise J. H., 2018, MNRAS, 475, 4868

Correa Magnus L., Vasiliev E., 2022, MNRAS, 511, 2610

Creevey O. L., et al., 2017, A&A, 601, A67

D'Onghia E., Lake G., 2008, ApJL, 686, L61

D'Souza R., Bell E. F., 2018, MNRAS, 474, 5300

D'Souza R., Bell E. F., 2021, MNRAS, 504, 5270

D'Souza R., Bell E. F., 2022, MNRAS, 512, 739

Da Costa G. S., Armandroff T. E., 1990, AJ, 100, 162

Da Costa G. S., et al., 2019, MNRAS, 489, 5900

Danieli S., van Dokkum P., Merritt A., Abraham R., Zhang J., Karachentsev I. D., Makarova L. N., 2017, ApJ, 837, 136

Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270

De Silva G. M., et al., 2015, MNRAS, 449, 2604

Deason A. J., Belokurov V., Evans N. W., 2011, MNRAS, 411, 1480

Deason A. J., Wetzel A. R., Garrison-Kimmel S., Belokurov V., 2015, MNRAS, 453, 3568

Deason A. J., Mao Y.-Y., Wechsler R. H., 2016, The Astrophysical Journal, 821, 5

Deason A. J., et al., 2021, MNRAS, 501, 5964

Di Matteo P., Spite M., Haywood M., Bonifacio P., Gómez A., Spite F., Caffau E., 2020, A&A, 636, A115

Diemer B., Kravtsov A. V., 2014, ApJ, 789, 1

Dixon K. L., Iliev I. T., Gottlöber S., Yepes G., Knebe A., Libeskind N., Hoffman Y., 2018, MNRAS, 477, 867

Eadie G., Jurić M., 2019, ApJ, 875, 159

Eadie G. M., Springford A., Harris W. E., 2017, ApJ, 835, 167

Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, ApJ, 136, 748

El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, ApJ, 820, 131

El-Badry K., Wetzel A. R., Geha M., Quataert E., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2017, ApJ, 835, 193

El-Badry K., et al., 2018a, MNRAS, 473, 1930

El-Badry K., et al., 2018b, Monthly Notices of the Royal Astronomical Society, 480, 652

Erkal D., et al., 2018, MNRAS, 481, 3148

Escala I., et al., 2018, MNRAS, 474, 2194

Evrard A. E., et al., 2002, ApJ, 573, 7

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

Fattahi A., et al., 2020, MNRAS, 497, 4459

Faucher-Giguère C.-A., 2020, MNRAS, 493, 1614

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416

Fillingham S. P., et al., 2019, arXiv e-prints, p. arXiv:1906.04180

Fillmore J. A., Goldreich P., 1984, ApJ, 281, 1

Fitts A., et al., 2017, MNRAS, 471, 3547

Forero-Romero J. E., Hoffman Y., Yepes G., Gottlöber S., Piontek R., Klypin A., Steinmetz M., 2011, MNRAS, 417, 1434

Freeman K., Bland-Hawthorn J., 2002, AR&AA, 40, 487

Fritz T. K., Battaglia G., Pawlowski M. S., Kallivayalil N., van der Marel R., Sohn S. T., Brook C., Besla G., 2018a, A&A, 619, A103

Fritz T. K., Lokken M., Kallivayalil N., Wetzel A., Linden S. T., Zivick P., Tollerud E. J., 2018b, ApJ, 860, 164

Gaia Collaboration et al., 2018, A&A, 616, A1

Gallart C., et al., 2015, ApJL, 811, L18

Gallart C., Bernard E. J., Brook C. B., Ruiz-Lara T., Cassisi S., Hill V., Monelli M., 2019, Nature Astronomy, 3, 932–939

Gandhi P. J., Wetzel A., Hopkins P. F., Shappee B. J., Wheeler C., Faucher-Giguère C.-A., 2022, MNRAS, 516, 1941

Garavito-Camargo N., Besla G., Laporte C. F. P., Johnston K. V., Gómez F. A., Watkins L. L., 2019, ApJ, 884, 51

Garavito-Camargo N., Besla G., Laporte C. F. P., Price-Whelan A. M., Cunningham E. C., Johnston K. V., Weinberg M., Gómez F. A., 2021, ApJ, 919, 109

Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, MNRAS, 438, 2578

Garrison-Kimmel S., Bullock J. S., Boylan-Kolchin M., Bardwell E., 2017a, MNRAS, 464, 3108

Garrison-Kimmel S., et al., 2017b, MNRAS, 471, 1709

Garrison-Kimmel S., et al., 2018, MNRAS, 481, 4133

Garrison-Kimmel S., et al., 2019a, MNRAS, 487, 1380

Garrison-Kimmel S., et al., 2019b, MNRAS, 489, 4574

Ge Z. S., Bi S. L., Chen Y. Q., Li T. D., Zhao J. K., Liu K., Ferguson J. W., Wu Y. Q., 2016, ApJ, 833, 161

Geha M., et al., 2017, ApJ, 847, 4

Gilmore G., Reid N., 1983, MNRAS, 202, 1025

Gómez F. A., Besla G., Carpintero D. D., Villalobos Á., O'Shea B. W., Bell E. F., 2015, ApJ, 802, 128

Gott J. R. I., Turner E. L., Aarseth S. J., 1979, ApJ, 234, 13

Grand R. J. J., Springel V., Gómez F. A., Marinacci F., Pakmor R., Campbell D. J. R., Jenkins A., 2016, MNRAS, 459, 199

Grand R. J. J., et al., 2017, MNRAS, 467, 179

Graus A. S., Bullock J. S., Boylan-Kolchin M., Weisz D. R., 2016, MNRAS, 456, 477

Graziani L., Salvadori S., Schneider R., Kawata D., de Bennassuti M., Maselli A., 2015, MNRAS, 449, 3137

Graziani L., de Bennassuti M., Schneider R., Kawata D., Salvadori S., 2017, MNRAS, 469, 1101

Griffen B. F., Ji A. P., Dooley G. A., Gómez F. A., Vogelsberger M., O'Shea B. W., Frebel A., 2016, ApJ, 818, 10

Griffen B. F., Dooley G. A., Ji A. P., O'Shea B. W., Gómez F. A., Frebel A., 2018, MNRAS, 474, 443

Grillmair C. J., Dionatos O., 2006, ApJL, 643, L17

Guedes J., Callegari S., Madau P., Mayer L., 2011, ApJ, 742, 76

Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1

Gurvich A. B., et al., 2022, arXiv e-prints, p. arXiv:2203.04321

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hawkins K., Jofré P., Gilmore G., Masseron T., 2014, MNRAS, 445, 2575

Haywood M., Di Matteo P., Lehnert M. D., Snaith O., Khoperskov S., Gómez A., 2018, ApJ, 863, 113

Helmi A., 2008, Astronomy and Astrophysics Review, 15, 145

Helmi A., White S. D. M., 2001, MNRAS, 323, 529

Helmi A., White S. D. M., de Zeeuw P. T., Zhao H., 1999, Nature, 402, 53

Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85

Hill A. R., Muzzin A., Franx M., Marchesini D., 2017, ApJL, 849, L26

Hopkins P. F., 2009, in Jogee S., Marinova I., Hao L., Blanc G. A., eds, Astronomical Society of the Pacific Conference Series Vol. 419, Galaxy Evolution: Emerging Insights and Future Challenges. p. 228

Hopkins P. F., 2015, MNRAS, 450, 53

Hopkins P. F., 2016, MNRAS, 455, 89

Hopkins P. F., Hernquist L., Martini P., Cox T. J., Robertson B., Di Matteo T., Springel V., 2005, ApJL, 625, L71

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hopkins P. F., et al., 2018, MNRAS, 480, 800

Hunter J. D., 2007, Computing in Science Engineering, 9, 90

Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194

Ibata R. A., Wyse R. F. G., Gilmore G., Irwin M. J., Suntzeff N. B., 1997, AJ, 113, 634

Ishchenko M., Sobolenko M., Berczik P., Khoperskov S., Omarov C., Sobodar O., Makukov M., 2023, arXiv e-prints, p. arXiv:2304.03547

Jahn E. D., Sales L. V., Wetzel A., Boylan-Kolchin M., Chan T. K., El-Badry K., Lazar A., Bullock J. S., 2019, MNRAS, 489, 5348

Jahn E. D., Sales L. V., Wetzel A., Samuel J., El-Badry K., Boylan-Kolchin M., Bullock J. S., 2022, MNRAS, 513, 2673

Jethwa P., Erkal D., Belokurov V., 2016, MNRAS, 461, 2212

Ji A. P., et al., 2021, ApJ, 921, 32

Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, ApJ, 675, 1095

Jiang F., Dekel A., Freundlich J., van den Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2021, MNRAS, 502, 621

Johnston K. V., Spergel D. N., Hernquist L., 1995, ApJ, 451, 598

Johnston K. V., Bullock J. S., Sharma S., Font A., Robertson B. E., Leitner S. N., 2008, ApJ, 689, 936

Kallivayalil N., van der Marel R. P., Alcock C., Axelrod T., Cook K. H., Drake A. J., Geha M., 2006, ApJ, 638, 772

Kallivayalil N., van der Marel R. P., Besla G., Anderson J., Alcock C., 2013, ApJ, 764, 161

Kallivayalil N., et al., 2018, ApJ, 867, 19

Karachentsev I. D., Kudrya Y. N., 2014, AJ, 148, 50

Kelley T., Bullock J. S., Garrison-Kimmel S., Boylan-Kolchin M., Pawlowski M. S., Graus A. S., 2019, MNRAS, 487, 4409

Kim J.-h., Wise J. H., Abel T., 2009, ApJL, 694, L123

Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999, ApJ, 516, 530

Koh D., Wise J. H., 2018, MNRAS, 474, 3817

Kourkchi E., Tully R. B., 2017, ApJ, 843, 16

Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73

Kroupa P., 2001, MNRAS, 322, 231

Kruijssen J. M. D., Pfeffer J. L., Reina-Campos M., Crain R. A., Bastian N., 2019, MNRAS, 486, 3180

Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36

Leavitt H. S., 1908, Annals of Harvard College Observatory, 60, 87

Lee M. G., Freedman W. L., Madore B. F., 1993, ApJ, 417, 553

Leitherer C., et al., 1999, The Astrophysical Journal Supplement Series, 123, 3

Leitner S. N., 2012, ApJ, 745, 149

Li Y.-S., Helmi A., 2008, MNRAS, 385, 1365

Li H.-N., Zhao G., Christlieb N., Wang L., Wang W., Zhang Y., Hou Y., Yuan H., 2015, ApJ, 798, 110

Li H., Tan K., Zhao G., 2018, ApJS, 238, 16

Li T. S., et al., 2019, MNRAS, 490, 3508

Li Z.-Z., Qian Y.-Z., Han J., Li T. S., Wang W., Jing Y. P., 2020a, ApJ, 894, 10

Li Z.-Z., Zhao D.-H., Jing Y. P., Han J., Dong F.-Y., 2020b, ApJ, 905, 177

Li T. S., et al., 2021, ApJ, 911, 149

Li T. S., et al., 2022, ApJ, 928, 30

Lin L., et al., 2008, ApJ, 681, 232

Lin L., et al., 2010, ApJ, 718, 1158

Loebman S. R., Roškar R., Debattista V. P., Ivezić Ž., Quinn T. R., Wadsley J., 2011, ApJ, 737, 8

Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, ApJ, 692, 931

Lunnan R., Vogelsberger M., Frebel A., Hernquist L., Lidz A., Boylan-Kolchin M., 2012, ApJ, 746, 109

Ma X., Hopkins P. F., Wetzel A. R., Kirby E. N., Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Quataert E., 2017, MNRAS, 467, 2430

Ma X., et al., 2018, MNRAS, 478, 1694

Magg M., Hartwig T., Agarwal B., Frebel A., Glover S. C. O., Griffen B. F., Klessen R. S., 2018, MNRAS, 473, 5308

Majewski S. R., Munn J. A., Hawley S. L., 1996, ApJL, 459, L73

Majewski S. R., Skrutskie M. F., Weinberg M. D., Ostheimer J. C., 2003, ApJ, 599, 1082

Majewski S. R., et al., 2017, AJ, 154, 94

Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, ApJ, 907, 85

Martig M., et al., 2016, MNRAS, 456, 3655

Martínez-Delgado D., Karim N., Charles E. J. E., Boschin W., Monelli M., Collins M. L. M., Donatiello G., Alfaro E. J., 2022, MNRAS, 509, 16

Matsuno T., Aoki W., Suda T., 2019, ApJL, 874, L35

McCluskey F., Wetzel A., Loebman S. R., Moreno J., Faucher-Giguere C.-A., 2023, arXiv e-prints, p. arXiv:2303.14210

McConnachie A. W., 2012, AJ, 144, 4

McMillan P. J., 2011, MNRAS, 414, 2446

McMillan P. J., 2017, MNRAS, 465, 76

Miglio A., et al., 2017, Astronomische Nachrichten, 338, 644

Miller T. B., van den Bosch F. C., Green S. B., Ogiya G., 2020, MNRAS, 495, 4496

Miyoshi T., Chiba M., 2020, ApJ, 905, 109

Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319

Monachesi A., et al., 2019, MNRAS, 485, 2589

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Müller O., Rejkuba M., Pawlowski M. S., Ibata R., Lelli F., Hilker M., Jerjen H., 2019, A&A, 629, A18

Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019, MNRAS, 488, 1235

Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, ApJ, 901, 48

Naidu R. P., et al., 2021, arXiv e-prints, p. arXiv:2103.03251

Navarro J. F., White S. D. M., 1994, MNRAS, 267, 401

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Necib L., Lisanti M., Garrison-Kimmel S., Wetzel A., Sanderson R., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2018, arXiv e-prints, p. arXiv:1810.12301

Necib L., et al., 2019, arXiv e-prints, p. arXiv:1907.07190

Newberg H. J., 2016, in Newberg H. J., Carlin J. L., eds, Astrophysics and Space Science Library Vol. 420, Tidal Streams in the Local Group and Beyond. p. 1 (arXiv:2103.04440), doi:10.1007/978-3-319-19336-6˙1

Newberg H. J., et al., 2002, ApJ, 569, 245

Newberg H. J., et al., 2003, ApJL, 596, L191

Newberg H. J., Yanny B., Willett B. A., 2009, ApJL, 700, L61

Nikakhtar F., et al., 2021, arXiv e-prints, p. arXiv:2104.08394

Nissen P. E., Schuster W. J., 2010, A&A, 511, L10

Nogueras-Lara F., et al., 2019, arXiv e-prints, p. arXiv:1910.06968

Norris J. E., Ryan S. G., 1989, ApJL, 336, L17

Oñorbe J., Garrison-Kimmel S., Maller A. H., Bullock J. S., Rocha M., Hahn O., 2014, MNRAS, 437, 1894

Oñorbe J., Hennawi J. F., Lukić Z., 2017, ApJ, 837, 106

Odenkirchen M., et al., 2001, ApJL, 548, L165

Ogiya G., Taylor J. E., Hudson M. J., 2021, MNRAS, 503, 1233

Oluseyi H. M., et al., 2012, AJ, 144, 9

Ostriker J. P., Tremaine S. D., 1975, ApJL, 202, L113

Pace A. B., Erkal D., Li T. S., 2022, arXiv e-prints, p. arXiv:2205.05699

Panithanpaisal N., Sanderson R. E., Wetzel A., Cunningham E. C., Bailin J., Faucher-Giguère C.-A., 2021, ApJ, 920, 10

Panithanpaisal N., Sanderson R. E., Arora A., Cunningham E. C., Baptista J., 2022, arXiv e-prints, p. arXiv:2210.14983

Papovich C., et al., 2015, ApJ, 803, 26

Pardy S. A., et al., 2020, MNRAS, 492, 1543

Patel E., Mandel K. S., 2022, arXiv e-prints, p. arXiv:2211.15928

Patel E., Besla G., Sohn S. T., 2017, MNRAS, 464, 3825

Patel E., Besla G., Mandel K., Sohn S. T., 2018, ApJ, 857, 78

Patel E., et al., 2020, ApJ, 893, 121

Pawlowski M. S., Tony Sohn S., 2021, ApJ, 923, 42

Peñarrubia J., Benson A. J., 2005, MNRAS, 364, 977

Peñarrubia J., Kroupa P., Boily C. M., 2002, MNRAS, 333, 779

Pearson W. J., et al., 2019, A&A, 631, A51

Peng E. W., Ford H. C., Freeman K. C., White R. L., 2002, AJ, 124, 3144

Peschken N., Athanassoula E., Rodionov S. A., 2017, MNRAS, 468, 994

Peschken N., Lokas E. L., Athanassoula E., 2020, MNRAS, 493, 1375

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06209

Press W. H., Schechter P., 1974, ApJ, 187, 425

Price-Whelan A. M., 2017, The Journal of Open Source Software, 2

Price-Whelan A. M., Johnston K. V., Valluri M., Pearson S., Küpper A. H. W., Hogg D. W., 2016, MNRAS, 455, 1079

Price-Whelan A. M., Mateu C., Iorio G., Pearson S., Bonaca A., Belokurov V., 2019, AJ, 158, 223

Putman M. E., Zheng Y., Price-Whelan A. M., Grcevich J., Johnson A. C., Tollerud E., Peek J. E. G., 2021, ApJ, 913, 53

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Renzini A., et al., 2018, ApJ, 863, 16

Riebe K., et al., 2013, Astronomische Nachrichten, 334, 691

Robertson B. E., 2022, AR&AA, 60, 121

Robertson B., Bullock J. S., Cox T. J., Di Matteo T., Hernquist L., Springel V., Yoshida N., 2006, ApJ, 645, 986

Robles V. H., Bullock J. S., 2021, MNRAS, 503, 5232

Rocha M., Peter A. H. G., Bullock J., 2012, MNRAS, 425, 231

Rodionov S. A., Athanassoula E., Peschken N., 2017, A&A, 600, A25

Rodriguez Wimberly M. K., Cooper M. C., Fillingham S. P., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., 2019, MNRAS, 483, 4031

Rodriguez Wimberly M. K., et al., 2022, MNRAS, 513, 4968

Safarzadeh M., Ji A. P., Dooley G. A., Frebel A., Scannapieco E., Gómez F. A., O'Shea B. W., 2018, MNRAS, 476, 5006

Sales L. V., Navarro J. F., Cooper A. P., White S. D. M., Frenk C. S., Helmi A., 2011, MNRAS, 418, 648

Sales L. V., Navarro J. F., Kallivayalil N., Frenk C. S., 2017, MNRAS, 465, 1879

Sales L. V., Wetzel A., Fattahi A., 2022, Nature Astronomy, 6, 897

Salvadori S., Schneider R., Ferrara A., 2007, MNRAS, 381, 647

Samuel J., et al., 2020, MNRAS, 491, 1471

Samuel J., Wetzel A., Chapman S., Tollerud E., Hopkins P. F., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., 2021, MNRAS, 504, 1379

Samuel J., Pardasani B., Wetzel A., Santistevan I., Boylan-Kolchin M., Moreno J., Faucher-Giguere C.-A., 2022a, arXiv e-prints, p. arXiv:2212.07518

Samuel J., Wetzel A., Santistevan I., Tollerud E., Moreno J., Boylan-Kolchin M., Bailin J., Pardasani B., 2022b, MNRAS, 514, 5276

Sand D. J., et al., 2022, ApJL, 935, L17

Sanderson R. E., et al., 2018a, arXiv e-prints, p. arXiv:1806.10564

Sanderson R. E., et al., 2018b, ApJ, 869, 12

Sanderson R. E., et al., 2020, ApJS, 246, 6

Santistevan I. B., Wetzel A., El-Badry K., Bland-Hawthorn J., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., Benincasa S., 2020, MNRAS, 497, 747

Santistevan I. B., Wetzel A., Sanderson R. E., El-Badry K., Samuel J., Faucher-Giguère C.-A., 2021, MNRAS, 505, 921

Santistevan I. B., Wetzel A., Tollerud E., Sanderson R. E., Samuel J., 2023, MNRAS, 518, 1427

Sawala T., et al., 2016, MNRAS, 457, 1931

Scannapieco E., Kawata D., Brook C. B., Schneider R., Ferrara A., Gibson B. K., 2006, ApJ, 653, 285

Schaye J., et al., 2015, MNRAS, 446, 521

Schiavon R. P., et al., 2017, MNRAS, 465, 501

Schuster W. J., Moreno E., Nissen P. E., Pichardo B., 2012, A&A, 538, A21

Searle L., Zinn R., 1978, ApJ, 225, 357

Sestito F., et al., 2019, Monthly Notices of the Royal Astronomical Society, 484, 2166

Sestito F., et al., 2020, MNRAS, 497, L7

Sestito F., et al., 2021, MNRAS, 500, 3750

Shipp N., et al., 2018, ApJ, 862, 114

Shipp N., et al., 2019, ApJ, 885, 3

Shipp N., et al., 2021, ApJ, 923, 149

Shipp N., et al., 2022, arXiv e-prints, p. arXiv:2208.02255

Silva Aguirre V., et al., 2018, MNRAS, 475, 5487

Simon J. D., 2019, AR&AA, 57, 375

Slater C. T., Bell E. F., 2013, ApJ, 773, 17

Smercina A., Bell E. F., Price P. A., D'Souza R., Slater C. T., Bailin J., Monachesi A., Nidever D., 2018, ApJ, 863, 152

Sohn S. T., Patel E., Fardal M. A., Besla G., van der Marel R. P., Geha M., Guhathakurta P., 2020, ApJ, 901, 43

Song M., et al., 2016, ApJ, 825, 5

Sparre M., Springel V., 2017, MNRAS, 470, 3946

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010a, AR&AA, 48, 391

Springel V., 2010b, MNRAS, 401, 791

Springel V., Hernquist L., 2005, ApJL, 622, L9

Springel V., et al., 2005, Nature, 435, 629

Starkenburg E., Oman K. A., Navarro J. F., Crain R. A., Fattahi A., Frenk C. S., Sawala T., Schaye J., 2017a, MNRAS, 465, 2212

Starkenburg E., et al., 2017b, MNRAS, 471, 2587

Steinmetz M., et al., 2006, AJ, 132, 1645

Stewart K. R., Bullock J. S., Wechsler R. H., Maller A. H., Zentner A. R., 2008, ApJ, 683, 597

Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, MNRAS, 471, 144

Tamfal T., Mayer L., Quinn T. R., Capelo P. R., Kazantzidis S., Babul A., Potter D., 2021, ApJ, 916, 55

Tapia T., Eliche-Moral M. C., Aceves H., Rodríguez-Pérez C., Borlaff A., Querejeta M., 2017, A&A, 604, A105

Taylor J. E., Babul A., 2001, ApJ, 559, 716

Teyssier R., 2002, A&A, 385, 337

Ting Y.-S., Conroy C., Goodman A., 2015, ApJ, 807, 104

Toomre A., Toomre J., 1972, ApJ, 178, 623

Torrealba G., Koposov S. E., Belokurov V., Irwin M., 2016, MNRAS, 459, 2370

Torrealba G., et al., 2019, MNRAS, 488, 2743

Valenti E., Zoccali M., Renzini A., Brown T. M., Gonzalez O. A., Minniti D., Debattista V. P., Mayer L., 2013, A&A, 559, A98

Vasiliev E., 2019a, MNRAS, 482, 1525

Vasiliev E., 2019b, MNRAS, 482, 1525

Vasiliev E., 2023, arXiv e-prints, p. arXiv:2304.09136

Vasiliev E., Belokurov V., Erkal D., 2021, MNRAS, 501, 2279

Venn K. A., et al., 2020, MNRAS, 492, 3241

Villalobos Á., Helmi A., 2008, MNRAS, 391, 1806

Virtanen P., et al., 2020, Nature Methods, 17, 261

Vogelsberger M., Mohayaee R., White S. D. M., 2011, MNRAS, 414, 3044

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Walker I. R., Mihos J. C., Hernquist L., 1996, ApJ, 460, 121

Warfield J. T., et al., 2023, MNRAS, 519, 1189

Webb J. J., Bovy J., 2020, MNRAS, 499, 116

Weinberg M. D., 1986, ApJ, 300, 93

Weinberg M. D., 1989, MNRAS, 239, 549

Weisz D. R., Boylan-Kolchin M., 2017, MNRAS, 469, L83

Weisz D. R., et al., 2023, arXiv e-prints, p. arXiv:2301.04659

Wetzel A., Garrison-Kimmel S., 2020a, HaloAnalysis: Read and analyze halo catalogs and merger trees (ascl:2002.014)

Wetzel A., Garrison-Kimmel S., 2020b, GizmoAnalysis: Read and analyze Gizmo simulations (ascl:2002.015)

Wetzel A. R., White M., 2010, MNRAS, 403, 1072

Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2014, MNRAS, 439, 2687

Wetzel A. R., Deason A. J., Garrison-Kimmel S., 2015, ApJ, 807, 49

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJL, 827, L23

Wheeler C., et al., 2019, MNRAS, 490, 4447

White S. D. M., 1976, MNRAS, 177, 717

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Yu S., et al., 2020, MNRAS, 494, 1539

Yu S., et al., 2021, arXiv e-prints, p. arXiv:2103.03888

Yuan Z., et al., 2019, arXiv e-prints, p. arXiv:1910.07538

Zhu H., Avestruz C., Gnedin N. Y., 2019, ApJ, 882, 152

Zinn R., 1993, in Smith G. H., Brodie J. P., eds, Astronomical Society of the Pacific Conference Series Vol. 48, The Globular Cluster-Galaxy Connection. p. 38

Zoccali M., et al., 2003, A&A, 399, 931

Zucker D. B., et al., 2004, ApJL, 612, L117

de Bennassuti M., Schneider R., Valiante R., Salvadori S., 2014, MNRAS, 445, 3039

de Bennassuti M., Salvadori S., Schneider R., Valiante R., Omukai K., 2017, MNRAS, 465, 926

del Pino A., et al., 2022, ApJ, 933, 76

van Dokkum P. G., et al., 2013, ApJL, 771, L35

van den Bosch F. C., Ogiya G., 2018, MNRAS, 475, 4066

van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, MNRAS, 387, 79

van den Bosch F. C., Ogiya G., Hahn O., Burkert A., 2018, MNRAS, 474, 3043

van der Marel R. P., Alves D. R., Hardy E., Suntzeff N. B., 2002, AJ, 124, 2639

van der Marel R. P., Fardal M., Besla G., Beaton R. L., Sohn S. T., Anderson J., Brown T., Guhathakurta P., 2012, ApJ, 753, 8

van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science Engineering, 13, 22