Title
Local Group Satellite Galaxies in Cosmological Simulations

Author
Samuel, Jenna

Publication Date
2021

Peer reviewed|Thesis/dissertation

Local Group Satellite Galaxies in Cosmological Simulations

By

JENNA MAE SAMUEL

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:

Professor Andrew Wetzel

Professor Christopher Fassnacht

Professor Lori Lubin

Committee in Charge

2021

In memory of my father.

Contents

Abstract

Satellite dwarf galaxies are singularly useful for studies of galaxy formation because they are highly susceptible to environmental and stellar feedback effects, we can observe them in unparalleled detail in the Local Group (LG), and we can model them at high resolution in simulations. Of particular interest in galaxy formation are the Milky Way's (MW) satellites, which appear to be nearly uniformly quenched inside the MW's virial radius, likely because of both internal stellar feedback and environmental effects of the gaseous host halo. Many observational campaigns have been devoted to measuring LG satellite star formation histories and high precision proper motions from resolved stellar populations, critical to determining orbital effects on formation. However, LG surveys will revolutionize galaxy formation, only if we can provide sufficiently rigorous theoretical models to interpret them.

The spatial distribution, kinematics, and star formation histories of satellite dwarf galaxies in the LG present a unique opportunity to test galaxy formation and numerical resolution in cosmological hydrodynamic simulations. We find excellent agreement between LG observations and the inner radial distributions of 'classical' dwarf satellites around MW/M31-mass host galaxies from the FIRE simulations. These results also suggest potentially undiscovered classical dwarf-mass satellites in the outer halos of the MW and M31. Furthermore, we investigate the prevalence, longevity, and potential causes of so-called 'satellite planes' in the LG. Although satellites in the simulations are typically distributed almost isotropically around their hosts, we find that hosts with an LMC-like satellite are more likely to have MW-like satellite planes, and that M31's satellite distribution is much more common. Preliminary results on the gas content and star formation histories of simulated satellites indicate agreement with the high quenched fraction of satellites in the LG. However, the simulations and LG observations appear to be in tension with the small fraction of quenched satellites around MW analogs in the nearby universe. Our results suggest that environmental effects of the host halo may be the dominant mechanisms for quenching satellite dwarf galaxies.

Acknowledgments

It has truly taken the support of a whole community to carry out the work culminating in this dissertation. Below, I thank a small sample of those who have helped make it possible.

To my advisor, Andrew Wetzel, you have been such an incredible force in my career, even in just a few years. Your commitment to your students and your science is an inspiration. Thank you for listening to me when I needed a dose of realism, and for pushing me when you knew I could succeed.

To my professors and collaborators, your feedback on my work and the examples you set have shaped the way I think about and do science. Thank you for welcoming me into the fold and showing me the ropes.

To my group mates, it has been a privilege to work with you. To Sarah Loebman and Samantha Benincasa, thank you for sharing your postdoc wisdom with care, whether near or far. To my fellow graduate students, it has been a pleasure to have such wonderful peers to laugh and learn with. And thank you to Theodore for being the best Chief Morale Officer.

To my friends and colleagues in the department, thank you for the science discussions, for always listening to my rants, and for grabbing probably way too many coffees with me. To my office mates of the past few years, Jessie Hirtenstein, Azalee Bostroem, Geoff Chen, Victoria Strait, Isaiah Santistevan, Morgane Konig, and Rose Baunach, thank you for being my little family at work.

To my husband, Chris, your constant reminders of how far I had come at every step and your pride in what I had already accomplished carried me through my most difficult times in graduate school. Thank you to my cats, who have comforted me through long nights and early mornings. To Raven, thank you for carrying on being my closest friend even after I moved across the country to follow my dreams. To my mom, my in laws, and my friends and family, thank you for always asking what it is I do and supporting me, regardless of how well I explain my work.

CHAPTER 1

Introduction

In the last century, astronomers discovered that the stars and gas in galaxies were moving as if unseen matter was gravitationally interacting with them (Zwicky, 1937; Rubin & Ford, 1970), and that distant galaxies were accelerating away from us as if acted upon by an unknown force (Riess et al., 1998; Perlmutter et al., 1999). While astronomers investigated these mysteries, cosmologists were working on the question of how the universe began. All of these topics converged with the discovery of a practically isotropic background radiation field, the cosmic microwave background (CMB) (Penzias & Wilson, 1965). Cosmologists quickly realized that the CMB is a relic of an earlier time in the history of the universe, when matter and energy were concentrated in a homogeneous mixture. The CMB lent credence to the idea that the universe began with the Big Bang, expanding from a singularity in an instant (Dicke et al., 1965). Eventually, a more detailed picture of the CMB revealed that the early universe was not completely homogeneous, and small regions were measured to be slightly hotter or cooler than the average temperature of ~ 2.7 K, indicative of density perturbations due to what we now call dark matter (e.g., Spergel et al., 2003; Planck Collaboration et al., 2018).

As the universe continued to evolve post-recombination (when matter and energy decoupled, subatomic particles combined into atoms, and the CMB was emitted), the regions of dark matter overdensity remained and became enhanced over time as more matter was pulled into their gravitational potential wells (Press & Schechter, 1974). We call these seeds of structure formation dark matter halos, and they form the backbone for the formation of galaxies through the cooling and collapsing of gas into stars (Searle & Zinn, 1978). Dark

matter halos started out relatively small and grew through mergers with each other (Ostriker & Tremaine, 1975), into larger halos (and large scale structures like cosmic sheets and filaments). Today, we observe the remnants of hierarchical structure formation as the galaxies we can observe – smaller galaxies are more numerous and larger galaxies are more rare, having cannibalized an entire population of smaller ones in order to form and grow.

Similar to the formation of massive galaxies from smaller ones, individual galaxies can also form gravitationally bound associations with each other called galaxy groups or clusters. Groups typically contain $\lesssim 100$ galaxies and have total masses around $10^{13}\,\mathrm{M}_\odot$, while clusters are more massive with masses in the range of $10^{14-15}\,\mathrm{M}_\odot$ and can contain ~ 1000 galaxies (Kravtsov & Borgani, 2012). The group and cluster environments can amass large concentrations of both baryonic matter and dark matter, which has profound effects on the evolution of galaxies within such crowded environments.

Our Galaxy, the Milky Way, is part of a small group of galaxies known as the Local Group. The Local Group (LG) is dominated by two galaxies of similar mass ($\sim 10^{12}\,\mathrm{M}_\odot$) and morphology, the Milky Way (MW) and Andromeda (M31), along with ~ 100 smaller galaxies within $\sim 1-2$ Mpc of the MW-M31 pair (McConnachie, 2012). However, some of the smaller galaxies in the LG are actually quite massive. For example, Triangulum (M33) near M31 has about 1/3 the stellar mass of the MW or M31, and the Large Magellanic Cloud (LMC) near the MW is about 10 times less massive than the MW or M31. Most of the other galaxies in the LG are less massive dwarf galaxies, and they have stellar masses around $10^{5-9}\,\mathrm{M}_\odot$.

Dwarf galaxies represent extremes of galaxy formation: ultrafaints ($M_* \lesssim 10^5\,\mathrm{M}_\odot$) are likely quenched by reionization (Bullock et al., 2000; Weisz et al., 2014; Rodriguez Wimberly et al., 2019), some experience bursts of star formation that drive gas out and shallow their inner density profiles (Pontzen & Governato, 2012; Oñorbe et al., 2015), and others are torn apart by their host galaxy's environment (Ibata et al., 1994). Of all the different mass

regimes ($M_* \sim 10^{2-12}$) of galaxies that we observe, dwarf galaxies are the most dark matter-dominated systems, making them some of the best astrophysical objects for small-scale tests of cosmology (Bullock & Boylan-Kolchin, 2017). The dwarf galaxies within the virial radii of the MW or M31 ($\sim 300 - 400$ kpc) are considered satellite galaxies. Within the host's virial radius, the densities of both dark matter and gas are greatly enhanced relative to the field, which has significant effects on the morphology, gas content, and star formation of satellites. Thus, satellite dwarf galaxies also present a unique opportunity to study effects of the host galaxy environment on the evolution of galaxies that are the most susceptible to effects like ram pressure and tidal stripping.

Part of why LG satellite dwarf galaxies are particularly compelling test beds for galaxy formation is because we can observe them in exquisite detail. Only in the LG do we simultaneously have observational access to HI gas, resolved stellar populations, and 6D phase space coordinates. Over 1,000 orbits of Hubble Space Telescope (HST) data have been dedicated to studying LG satellites, including two recent Treasury Programs (GO—14734, PI Kallivayalil, 164 orbits; GO-15902, PI Weisz, 244 orbits) that are surveying the entire known satellite galaxy population of both the MW and M31 to provide first epoch proper motion measurements. Gaia has recently discovered faint MW satellites, and provided even higher precision proper motions to model satellite dynamics (e.g., Fritz et al., 2018; Kallivayalil et al., 2018). To place the MW in context of a statistical sample of similar galaxies, the Satellites Around Galactic Analogs (SAGA) survey targets satellite galaxies of MW analogs within 20-40 Mpc of the LG (Geha et al., 2017). The Pan-Andromeda Archaeological Survey (PAndAS) has also thoroughly imaged a vast region around M31 that encompasses many of its satellite galaxies, giving insight into their stellar populations and line-of-sight distances and velocities (e.g., Conn et al., 2012; Martin et al., 2013b). Current surveys are exploring the satellite populations around ~ 10 nearby MW analogs, using techniques like surface

brightness fluctuations to determine satellite distances (Smercina et al., 2021). Future observations with HST, Gaia, James Webb Space Telescope (JWST), Vera C. Rubin Observatory, and the Nancy Grace Roman Space Telescope (NGRST) are slated to further revolutionize our understanding of LG satellites and galaxy formation more generally.

Throughout the last several decades, observations of the LG have been instrumental in rigorously testing small-scale predictions from ΛCDM cosmology. While the Magellanic Clouds are visible with the naked eye from the southern hemisphere, fainter MW satellite dwarf galaxies were only discovered using telescopes starting in the mid-1900's. The Sloan Digital Sky Survey (SDSS) and Hubble Space Telescope (HST) pushed dwarf discovery even further, imaging down to the faint magnitudes necessary to detect and study the diffuse stellar distributions of nearby dwarfs (Belokurov et al., 2007; Brown et al., 2012).

As more satellites were discovered, certain peculiarities of LG (and especially MW) satellites with respect to dark matter-only (DMO) simulations became apparent. For example, DMO simulations predicted that there should be thousands of subhalos capable of hosting luminous dwarf galaxies around a MW/M31-mass host halo, but even as of today the total number of observed dwarfs in the LG is only ~ 100 (Moore et al., 1999; Klypin et al., 1999b). This mismatch in the predicted and observed number of satellite dwarf galaxies led to one of the most famous small-scale tensions with ΛCDM, the 'missing satellites problem'. While the missing satellites problem was a looming issue for galaxy formation simulations for almost two decades, the inclusion of baryonic physics in simulations has resolved the tension between simulations and observations in recent years (Kim et al., 2018; Buck et al., 2019; Garrison-Kimmel et al., 2019a).

Beyond just the discovery of LG dwarfs and their number counts, repeated imaging with HST and now Gaia has enabled the measurement of stellar proper motions, which can be combined with line-of-sight velocities from spectroscopy to yield full 3D motions of stars in LG dwarf galaxies. These proper motions have just begun to open up LG dwarf science to

a whole new perspective using the dynamics of stars and dwarfs to trace their evolutionary history. For example, 3D positions and velocities of satellite dwarfs around the MW have revealed that not only do they lie within a thin plane, but they also coherently orbit around the MW from within that plane (Lynden-Bell, 1976; Pawlowski & Kroupa, 2020). Such a 'satellite plane' may seem a bit unusual, and indeed it appears to be an exceedingly rare configuration in many (especially DMO) simulations (Metz et al., 2008; Pawlowski et al., 2014; Buck et al., 2016). Though a significant amount of work has gone into investigating the MW's satellite plane, there is currently no conclusive explanation of its origin.

The stellar populations and gas content of LG satellites have also been studied in detail, leading to the revelation that most LG satellites are gas-poor and no longer forming stars (Grcevich & Putman, 2009; Spekkens et al., 2014; Wetzel et al., 2015b; Putman et al., 2021). In contrast, isolated dwarf galaxies in the field are observed to be nearly completely star-forming (Geha et al., 2012). This juxtaposition is not too surprising given the inhospitable effects of the host galaxy environment like tidal disruption and ram pressure stripping, which can heat and remove the gas that fuels star formation and tear dwarfs apart. What is more interesting is that almost all satellite galaxies around nearby MW analogues from the SAGA survey appear to be star-forming (Geha et al., 2017; Mao et al., 2021). The underlying difference between LG and SAGA satellites that causes this tension remains an open question.

Recent computational advances combined with the small sizes and masses of dwarf galaxies also make it possible to model them at unprecedented resolution within a cosmological environment. The FIRE simulations have been able to push hydrodynamic resolution to unprecedented levels thanks to a new class of mesh-free Lagrangian hydrodynamics solvers. In particular, Wetzel et al. (2016) and Garrison-Kimmel et al. (2019a) have simulated suites of isolated MW-mass and paired LG-like systems. These suites, known as Latte and ELVIS on FIRE, resolve satellite dwarf galaxies down to the lower limits on classical dwarf mass

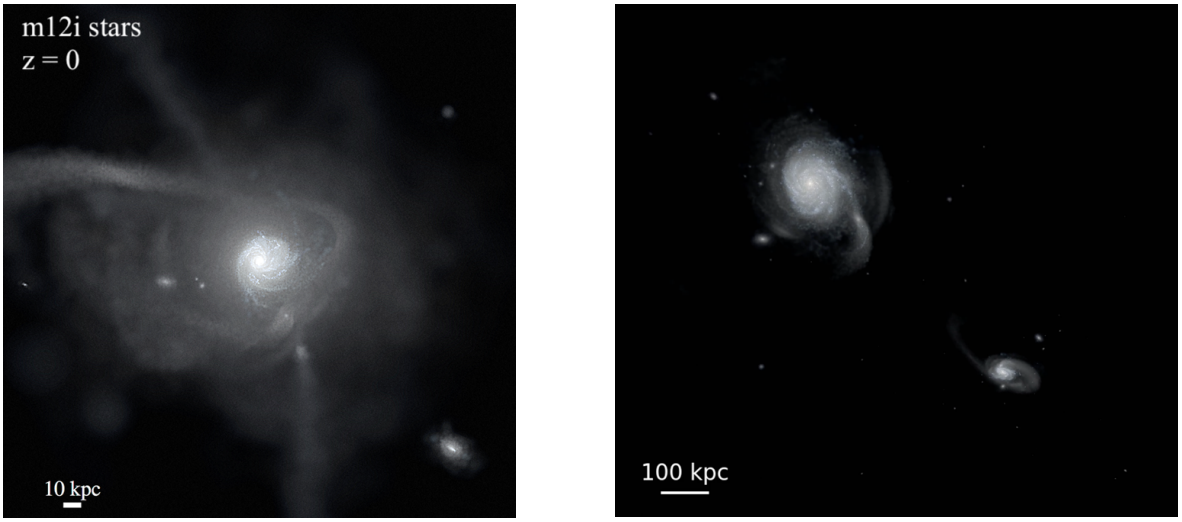

FIGURE 1.1. *Left:* The stellar distribution of a MW-mass galaxy simulation from FIRE. A prominent disk structure with spiral arms is surrounded by a diffuse stellar halo, satellite dwarf galaxies, and stellar streams. FIRE simulations have a cosmological environment consistent with observations of the LG. *Right:* Same as left, but for a pair of MW-mass galaxies in a configuration similar to the MW and M31 in the LG.

($M_* \sim 10^5 M_\odot$), and can therefore be used to study throughly the formation and evolution of satellites. Figure 1.1 shows images of the stars in Latte and ELVIS host galaxies, as well as their satellite dwarf galaxies.

Here I present my work on three studies of the spatial distribution, kinematics, and quenching of satellite dwarf galaxies in the FIRE simulations. Chapter 2 presents the radial distributions of satellite dwarfs around MW/M31-mass hosts in the FIRE simulations, including tests of physics and numerical resolution in simulations, the importance of the host disc in satellite disruption, and predictions for satellite discovery in the LG. Chapter 3 investigates the prevalence of satellite planes in these simulations, their lifetimes, observational considerations, and potential causes of satellite plane formation. Chapter 4 examines simulated satellite gas content, quenching in the host environment, and group pre-processing of satellites.

CHAPTER 2

A profile in FIRE: resolving the radial distributions of satellite galaxies in the Local Group with simulations

Published as Jenna Samuel, Andrew Wetzel, Erik Tollerud, Shea Garrison-Kimmel, Sarah Loebman, Kareem El-Badry, Philip F Hopkins, Michael Boylan-Kolchin, Claude-André Faucher-Giguère, James S Bullock, Samantha Benincasa, and Jeremy Bailin in Monthly Notices of the Royal Astronomical Society, Volume 491, Issue 1, January 2020, Pages 1471–1490, `https://doi.org/10.1093/mnras/stz3054`

2.1. Abstract

While many tensions between Local Group (LG) satellite galaxies and ΛCDM cosmology have been alleviated through recent cosmological simulations, the spatial distribution of satellites remains an important test of physical models and physical versus numerical disruption in simulations. Using the FIRE-2 cosmological zoom-in baryonic simulations, we examine the radial distributions of satellites with $M_* > 10^5\,M_\odot$ around 8 isolated Milky Way- (MW) mass host galaxies and 4 hosts in LG-like pairs. We demonstrate that these simulations resolve the survival and physical destruction of satellites with $M_* \gtrsim 10^5\,M_\odot$. The simulations broadly agree with LG observations, spanning the radial profiles around the MW and M31. This agreement does not depend strongly on satellite mass, even at distances $\lesssim$100 kpc. Host-to-host variation dominates the scatter in satellite counts within 300 kpc of the hosts, while time variation dominates scatter within 50 kpc. More massive host galaxies within our sample have fewer satellites at small distances, likely because of enhanced tidal destruction of satellites via the baryonic disks of host galaxies. Furthermore, we quantify

and provide fits to the tidal depletion of subhalos in baryonic relative to dark matter-only simulations as a function of distance. Our simulated profiles imply observational incompleteness in the LG even at $M_* \gtrsim 10^5\,M_\odot$: we predict 2-10 such satellites to be discovered around the MW and possibly 6-9 around M31. To provide cosmological context, we compare our results with the radial profiles of satellites around MW analogs in the SAGA survey, finding that our simulations are broadly consistent with most SAGA systems.

2.2. Introduction

Dark matter dominates the matter content of dwarf galaxies by up to several orders of magnitude, making them ideal sites for small-scale tests of the standard paradigm for structure formation: cold dark matter (CDM) with a cosmological constant (Λ). CDM makes testable predictions for both the central mass profile of dwarf galaxies and their number density and spatial distribution around more massive host galaxies. However, on such small scales, tests of CDM require highly resolved observations that are only feasible within the nearby Universe. Fortunately, the Milky Way (MW) and Andromeda (M31) galaxies that make up the Local Group (LG) are host to populations of satellite dwarf galaxies which can provide quantitative tests of CDM on small scales.

LG satellite galaxies have been a source of significant tensions within the CDM model, largely stemming from comparisons of observations to dark matter-only (DMO) simulations that lack the effects of baryonic physics. Arguably the most famous of these tensions, the "missing satellites" problem, describes a discrepancy between the number of subhalos predicted by DMO simulations compared to the smaller number of luminous satellite galaxies observed around the MW (e.g. Moore et al., 1999; Klypin et al., 1999b). However, newer simulations that include the effects of baryonic physics through hydrodynamics and sub-grid models show agreement with the number of observed satellite dwarf galaxies in the LG, in part from enhanced tidal disruption of satellites by the baryonic disks of host galaxies (e.g. Brooks et al., 2013; Sawala et al., 2016; Wetzel et al., 2016; Garrison-Kimmel et al.,

2018; Kelley et al., 2019; Simpson et al., 2018; Buck et al., 2019). N-body simulations in conjunction with semi-analytic models of galaxy formation have also yielded similar results, showing similar radial distributions for observable satellites (e.g. Macciò et al., 2010; Font et al., 2011). Simultaneously, a better understanding of observational incompleteness has also been critical in alleviating the missing satellites tension (Tollerud et al., 2008; Walsh et al., 2009; Hargis et al., 2014; Kim et al., 2018).

In addition to over-predicting the number of satellites, DMO simulations predict too many dense, massive ("too-big-to-fail") satellites (Boylan-Kolchin et al., 2011, 2012), and satellites with steeper ("cuspier") central density profiles than seen in observations (Navarro et al., 1996). Again, baryonic effects in simulations are a pathway to reconciling these problems because stellar feedback acts to redistribute the central dark matter and "core-out" the density profile of dwarfs (e.g. Mashchenko et al., 2008; Chan et al., 2015; Oñorbe et al., 2015; El-Badry et al., 2016; Dutton et al., 2016; Fitts et al., 2017).

Baryonic effects are also crucial for understanding the predicted phase space coordinates of satellites around simulated MW/M31-like galaxies. This phase space information can be used to infer the formation history of satellites and rigorously test CDM predictions. For example, orbit modeling of the Large Magellanic Cloud (LMC) has provided evidence that it is undergoing its first pericentric passage around the MW, and this may partly be why it is still able to form stars (Besla et al., 2007; Kallivayalil et al., 2013). The phase space distribution of LG satellites have further challenged the CDM model because the MW's satellite galaxies appear to be arranged in a thin, planar structure that is coherently rotating, and a similar structure has been found around M31 (Lynden-Bell, 1976; Metz et al., 2007; Conn et al., 2013; Ibata et al., 2013; Pawlowski, 2018).

Furthermore, the Satellites Around Galactic Analogs (SAGA) survey is broadening our understanding of LG satellites by targeting satellites of MW analogs within 20-40 Mpc of the LG (Geha et al., 2017). Their goal is to obtain a complete census of satellites around 100 MW

analogs, down to the luminosity of the Leo I dwarf galaxy ($M_r < -12.3$ or $M_* \approx 5 \times 10^6\,\mathrm{M}_\odot$). This will make it possible to connect LG satellite galaxies with a large sample of satellite populations, providing a statistically robust cosmological context to interpret LG galaxy formation and evolution.

Satellite dwarf galaxies can be used to study the effects of environment on galaxy formation as well. Even before satellites accrete onto their host, they are preprocessed by interactions with other dwarf galaxies that are bound to them in small groups (e.g. Zabludoff & Mulchaey, 1998; McGee et al., 2009; Wetzel et al., 2013; Hou et al., 2014). Once they fall into their host halo, satellites can be tidally disrupted into diffuse stellar structures by their host. For instance, the Sagittarius dwarf galaxy is being actively torn apart into a stellar stream within the MW's halo (e.g. Lynden-Bell & Lynden-Bell, 1995; Belokurov et al., 2006). As satellites orbit in the halos of their host galaxies, they are thought to be ram pressure-stripped of their gas, causing their star formation to be subsequently suppressed (e.g. Gunn & Gott, 1972; Fillingham et al., 2016). The MW and M31 may exert some of the strongest observed environmental effects on their satellite populations: most of their satellites are gas-poor and no longer forming stars, making them an interesting case study for environmental effects (e.g. Einasto et al., 1974; Mateo, 1998; Grcevich & Putman, 2009; McConnachie, 2012; Slater & Bell, 2014; Spekkens et al., 2014; Wetzel et al., 2015b).

Given the unique ability to measure full 3D positions and velocities of LG satellites, and thus infer their orbital histories, the LG also provides a fertile physical testing ground for numerical evolution and disruption of subhalos in simulations. Historically, it has been difficult to use simulations to interpret observations of LG satellites because baryonic simulations have only recently begun to produce dwarf galaxies that do not suffer from numerical over-merging. Simulations of satellites undergoing tidal disruption have revealed that the most critical simulation parameters in dynamically resolving satellites are spatial and mass resolution (e.g. Carlberg, 1994; van Kampen, 1995; Moore et al., 1996; Klypin et al., 1999a;

van Kampen, 2000; Diemand et al., 2007; Wetzel & White, 2010; van den Bosch & Ogiya, 2018). While current large-volume simulations can offer a sizable sample of satellite galaxies, their dwarf galaxies have limited resolution both in terms of particle mass and gravitational force softening, which curbs their usefulness in tests that require accurate tidal disruption and survival of satellites.

Understanding formation and evolution is contingent on resolving the radial distribution of satellites as a function of distance from their hosts in cosmological simulations. Higher resolution, 'zoom-in' simulations are now providing satellite populations that are sufficiently well-resolved for studying LG satellite populations in detail. The main questions this paper aims to answer are:

- Do cosmological zoom-in baryonic simulations reproduce the observed radial distributions of satellites around the MW, M31, and MW analogs?
- Do the radial profiles reflect physical disruption from the host galaxy and/or numerical disruption inherent in the simulations?
- How do radial profiles in hydrodynamic simulations differ from those in DMO simulations?
- If the simulations are representative of the LG, how complete are observations of dwarf galaxies out to large distances around the MW and M31?

This paper is organized as follows: in Section 2.3 we describe the simulations used and how satellites were selected from them, in Section 2.4 we describe the observational data set used, in Section 2.5 we present our results on radial profiles with comparisons of the hydrodynamic simulations to both observations and dark matter-only simulations, and a discussion of the conclusions and implications is given in Section 2.6.

2.3. Simulations

2.3.1. FIRE simulation suite. We use cosmological zoom-in baryonic simulations from the Feedback In Realistic Environments (FIRE) project[1], run with the upgraded FIRE-2 (Hopkins et al., 2018) numerical implementations of fluid dynamics, star formation, and stellar feedback. The FIRE-2 simulations use a Lagrangian meshless finite-mass (MFM) hydrodynamics code, `GIZMO` (Hopkins, 2015). The MFM method allows for hydrodynamic gas particle smoothing to adapt based on the density of particles while still conserving mass, energy, and momentum to machine accuracy. Gravitational forces are solved using an improved version of the N-body `GADGET-3` Tree-PM solver (Springel, 2005), and the gravitational force softening of gas particles automatically adapts to match their hydrodynamic smoothing length.

The FIRE-2 simulations invoke realistic gas physics through a metallicity-dependent treatment of radiative heating and cooling over $10 - 10^{10}$ K, including free-free, photo-ionization and recombination, Compton, photo-electric and dust collisional, cosmic ray, molecular, metal-line, and fine-structure processes, accounting for 11 elements (H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe). The cosmic UVB background is included using the Faucher-Giguère et al. (2009) model, in which HI reionization occurs early ($z_{\rm reion} \sim 10$). The simulations that we use also model sub-grid diffusion of metals via turbulence (Hopkins, 2016; Su et al., 2017; Escala et al., 2018).

Star formation occurs in gas that is self-gravitating, Jeans-unstable, cold ($T < 10^4$ K), dense ($n > 1000$ cm^{-3}), and molecular (following Krumholz & Gnedin 2011). Each star particle represents a single stellar population under the assumption of a Kroupa stellar initial mass function (Kroupa, 2001), and we evolve star particles according to standard stellar population models from `STARBURST99` v7.0 (Leitherer et al., 1999). The simulations explicitly

[1] `https://fire.northwestern.edu/`

model several stellar feedback processes including core-collapse and Type Ia supernovae, continuous stellar mass loss, photoionization, photoelectric heating, and radiation pressure.

For all simulations, we generate cosmological zoom-in initial conditions at $z = 99$ using the `MUSIC` code (Hahn & Abel, 2011), and we save 600 snapshots from z =99 to 0, with typical spacing of $\lesssim$25 Myr.

We use two suites of simulations in this paper. The first is the Latte suite of individual MW/M31-mass halos introduced in Wetzel et al. (2016). Latte consists of 7 hosts with halo masses $M_{200m} = 1 - 2 \times 10^{12}$ $M_\odot$ (where '200m' indicates a measurement relative to 200 times the mean matter density of the Universe), selected from a periodic volume of length 85.5 Mpc. Gas and star particles have initial masses of 7070 $M_\odot$, though because of stellar mass loss, at $z = 0$ a typical star particle has mass $\approx$ 5000 $M_\odot$. Dark matter particles have a mass resolution of $m_{dm} = 3.5 \times 10^4$ $M_\odot$. Dark matter and stars have fixed gravitational softening (comoving at $z > 9$ and physical at $z < 9$): $\epsilon_{dm} = 40$ pc and $\epsilon_{star} = 4$ pc (Plummer equivalent). The minimum gas resolution (inter-element spacing) and softening length reached in each simulation is ≈ 1 pc.

In this paper we introduce two new hosts into the Latte suite: m12w and m12r. We select them using the same criteria as the Latte suite: $M_{200m}(z = 0) = 1 - 2 \times 10^{12}$ $M_\odot$ and no neighboring halos of similar mass ($> 3 \times 10^{11}$ $M_\odot$) within at least 5 R_{200m}, to limit computational cost. However, for these two halos we impose an additional criterion: each must host an LMC-mass subhalo. Specifically, within the initial dark-matter-only simulation, we select halos that host (only) one subhalo within the following limits at $z = 0$: maximum circular velocity $V_{circ,max} = 92 \pm 12$ km/s, distance $d = 51 \pm 40$ kpc, radial velocity $v_{rad} = 64 \pm 17$ km/s, tangential velocity $v_{tan} = 314 \pm 60$ km/s. These criteria are centered on the observed values for the LMC (e.g., Kallivayalil et al., 2013; van der Marel & Kallivayalil, 2014), though we use a wider selection window than the observational uncertainties to find a sufficient sample in our cosmological volume, which for this sample is a periodic box of length

172 Mpc with updated cosmology to match Planck Collaboration et al. (2018): $h = 0.68$, $\Omega_\Lambda = 0.69$, $\Omega_m = 0.31$, $\Omega_b = 0.048$, $\sigma_8 = 0.82$, $n_s = 0.97$. The zoom-in re-simulations use the same resolution as the existing Latte suite (given the slightly different cosmology, dark-matter particles have slightly higher mass of $m_{dm} = 3.9 \times 10^4$ $M_\odot$). While we select these halos to have LMC-like subhalos in the pilot dark-matter-only simulation, when we re-run with baryonic physics the details of the satellite orbit (in particular the orbital phase) do change. m12w's most massive satellite has $M_* = 8 \times 10^8$ $M_\odot$ and at $z = 0$ is at $d = 248$ kpc, having experienced pericentric passage of 78 kpc 2.4 Gyr ago ($z = 0.19$). m12r's most massive satellite has $M_* = 2.8 \times 10^9$ $M_\odot$ and at $z = 0$ is at $d = 390$ kpc, having experienced pericentric passage of 30 kpc 0.7 Gyr ago at $z = 0.05$. We will examine the dynamics of these LMC-like passages in upcoming work (Chapman et al., in preparation).

In addition to the Latte suite, we include one additional individual host (m12z), selected to have a slightly lower halo mass at $z = 0$ and simulated at a higher mass resolution of $m_{baryon,ini} = 4200$ $M_\odot$ (Garrison-Kimmel et al., 2018).

We also use the ELVIS on FIRE suite of two simulations, which selected halos to mimic the separation and relative velocity of the MW-M31 pair in the LG (Garrison-Kimmel et al., 2018). These simulations have $\approx 2\times$ better mass resolution than the Latte suite: the Romeo & Juliet simulation has $m_{baryon,ini} = 3500$ $M_\odot$ and the Thelma & Louise simulation has $m_{baryon,ini} = 4000$ $M_\odot$.

All simulations assume flat ΛCDM cosmologies, with slightly different parameters across the full suite: $h = 0.68 - 0.71$, $\Omega_\Lambda = 0.69 - 0.734$, $\Omega_m = 0.266 - 0.31$, $\Omega_b = 0.0455 - 0.048$, $\sigma_8 = 0.801 - 0.82$, and $n_s = 0.961 - 0.97$, broadly consistent with Planck Collaboration et al. (2018).

2.3.2. Halo finder. We identify dark-matter (sub)halos using the `ROCKSTAR` 6D halo finder (Behroozi et al., 2013a). We identify halos according to their radius that encloses 200 times the mean matter density, R_{200m}, and keep those with bound mass fraction > 0.4 and

at least 30 dark matter particles. We generate a halo catalog at each of the 600 snapshots for each simulation. We then construct merger trees using `CONSISTENT-TREES` (Behroozi et al., 2013b). For numerical stability, we generate halo catalogs and merger trees using only dark matter particles.

We then assign star particles to each (sub)halo in post-processing as follows (adapted from the method described in Necib et al. 2018). Given each (sub)halo's radius, R_{halo}, and $V_{circ,max}$ as returned by `ROCKSTAR`, we first identify all star particles whose position is within 0.8 R_{halo} (out to a maximum radius of 30 kpc) and whose velocity is within 2 $V_{circ,max}$ of each (sub)halo's center-of-mass velocity. We then keep star particles (1) whose positions are within 1.5 R_{90} (the radius that encloses 90 per cent of the mass of member star particles) of both the center-of-mass position of member stars and the dark matter halo center (thus ensuring the galaxy center is coincident with the halo center), and (2) whose velocities are within 2 σ_{vel} (the velocity dispersion of member star particles) of the center-of-mass velocity of member stars. We then iteratively repeat (1) and (2) until M_*, the sum of the masses of all member star particles, converges to within 1 per cent. We keep all halos with at least 6 star particles and average stellar density $> 300\ M_\odot\,kpc^{-3}$.

We examined each galaxy in our sample at $z = 0$ by eye and found that this method robustly identifies real galaxies with stable properties across time; in particular, it reliably separates true galaxies from transient alignments between subhalos and stars in the stellar halos of the MW-mass hosts. All of the subhalos (within 300 kpc of their host) that we analyze are uncontaminated by low-resolution dark matter.

TABLE 2.1. Host galaxy properties and satellite galaxy counts

Name	M_{200m} [10^{12} $M_\odot$]	M_* [10^{10} $M_\odot$]	$N_{sat}(d < 50$ kpc)	$N_{sat}(d < 100$ kpc)	$N_{sat}(d < 300$ kpc)
MW	∼1.4	∼5	1 ± 0.5 (50%)	6 ± 0.5 (10%)	13 ± 0 (0%)
M31	∼1.6	∼10	2 ± 0.5 (25%)	5 ± 1.0 (20%)	27 ± 0.5 (¡5%)
m12m	1.6	10.0	1 ± 1.0 (100%)	7 ± 3.0 (45%)	27 ± 2.0 (10%)
m12b	1.4	7.3	0 ± 0.2 (N/A)	3 ± 1.0 (35%)	11 ± 0.7 (5%)
m12f	1.7	6.9	0 ± 0.0 (N/A)	1 ± 1.0 (100%)	16 ± 1.0 (5%)
Thelma	1.4	6.3	1 ± 0.5 (50%)	6 ± 1.2 (20%)	17 ± 1.2 (10%)
Romeo	1.3	5.9	1 ± 0.7 (70%)	4 ± 0.7 (20%)	17 ± 1.0 (5%)
m12i	1.2	5.5	0 ± 0.5 (N/A)	3 ± 1.2 (40%)	12 ± 0.5 (5%)
m12c	1.4	5.1	1 ± 0.9 (90%)	8 ± 1.5 (20%)	23 ± 1.0 (5%)
m12w	1.1	4.8	0 ± 0.5 (N/A)	5 ± 0.9 (20%)	22 ± 1.5 (10%)
Juliet	1.1	3.4	1 ± 0.5 (50%)	8 ± 1.7 (20%)	20 ± 0.5 (5%)
Louise	1.2	2.3	1 ± 0.5 (50%)	8 ± 1.5 (20%)	23 ± 0.7 (5%)
m12z	0.9	1.8	1 ± 0.5 (50%)	7 ± 0.7 (10%)	17 ± 0.7 (5%)
m12r	1.1	1.5	1 ± 1.2 (120%)	5 ± 1.7 (35%)	14 ± 2.2 (15%)
Time variation			1 ± 0.6 (60%)	5 ± 1.3 (25%)	17 ± 1.1 (5%)
Host-to-host variation			1 ± 0.5 (50%)	5 ± 2.5 (50%)	17 ± 4.7 (30%)
Total variation			1 ± 1.0 (100%)	5 ± 3.0 (60%)	17 ± 6.0 (35%)

(1) Name of the host. (2) Host halo mass (M_{200m}) at $z = 0$. The halo mass of the MW is calculated by taking the value of M_{200c} from Bland-Hawthorn & Gerhard (2016) and multiplying by the average value of M_{200m}/M_{200c} for the simulations. The halo mass of M31 is calculated similarly, using the value of M31's M_{200c} from van der Marel et al. (2012). (3) Host stellar mass at $z = 0$. The stellar mass of the MW is taken from Bland-Hawthorn & Gerhard (2016), and the stellar mass of M31 is taken from Sick et al. (2015). Simulated hosts are ordered from greatest to least stellar mass. (4-6) Median and scatter in the cumulative number of satellite galaxies with $M_* > 10^5$ $M_\odot$ at different distances from the host. For the MW and M31, the scatter is the 68 per cent variation from observational uncertainties from Figure 2.1. For the individual simulated hosts, the scatter is the 68 per cent variation over time from Figure 2.2, spanning 1.3 Gyr ($z = 0 - 0.1$ in steps of $z = 0.01$). The percentage in parentheses is the scatter normalized to the median number of satellites.

2.3.3. Satellite selection. We refer to the MW- and M31-mass galaxies in our simulations as "hosts" and their surrounding populations of dwarf galaxies with $M_* > 10^5$ $M_\odot$ within 300 kpc as "satellites". Each of the eight Latte+m12z simulations contains a single isolated host while each of the two ELVIS on FIRE simulations contains two hosts in a LG-like pair, with their own distinct satellite populations. This provides a total of 12 host-satellite systems to study and compare to observations. Table 2.1 summarizes properties of these systems. Host galaxies have stellar masses $M_* \sim 10^{10-11}$ $M_\odot$ and dark matter halos $M_h = 0.9 - 1.7 \times 10^{12}$ $M_\odot$. Host stellar mass is measured by computing the stellar mass enclosed by a 2D radius in the plane of the host disk and a height above and below the plane that together define a cylinder containing 90 per cent of the total stellar mass within a sphere of radius 30 kpc around the host galaxy.

Our satellite selection of $M_* > 10^5$ $M_\odot$ corresponds to a minimum of ∼20 star particles and a peak halo mass (throughout their history) of $M_{peak} > 8 \times 10^8$ $M_\odot$ (or $\gtrsim 2.3 \times 10^4$ dark matter particles prior to infall). We expect subhalos that contain satellite galaxies with $M_* \sim 10^5$ $M_\odot$ to be both resolved in the simulations (Hopkins et al., 2018) and nearly complete in observations (e.g. Koposov et al., 2007; Tollerud et al., 2008; Walsh et al., 2009; Tollerud et al., 2014; Martin et al., 2016), so we choose this as our lower mass limit to make reasonable comparisons to the MW and M31 (but see Sections 2.5.8 and 2.5.9 for further discussion on potential incompleteness in the LG). For this analysis, we consider only the total (3D) radial distance from satellite to host galaxy, leaving a complete study of the full 3D positions and the problem of satellite planes for future work. For further details on the stellar masses, velocity dispersions, dynamical masses, and star-formation histories of dwarf galaxies in our simulations, see Garrison-Kimmel et al. (2019a,b).

2.4. Observations

We use the compilation of observed stellar masses of LG satellite galaxies in Garrison-Kimmel et al. (2018), in which they assume stellar mass-to-light ratios from Woo et al. (2008)

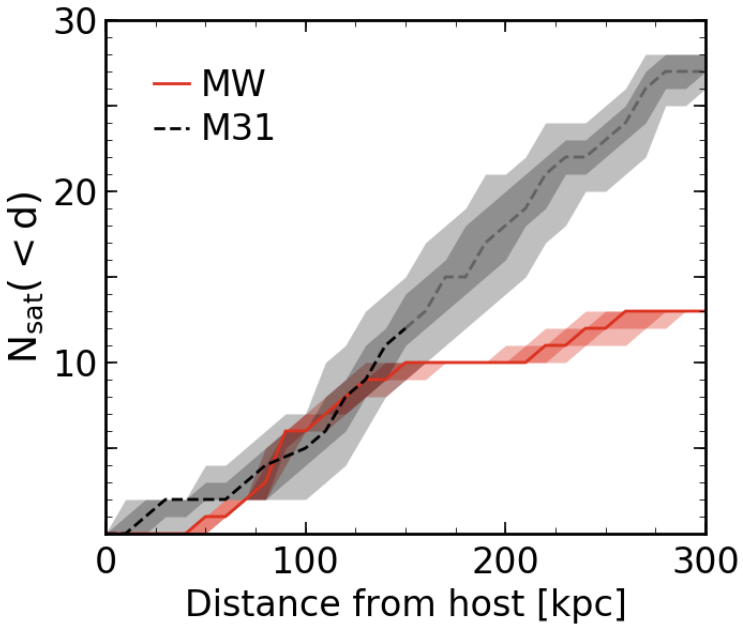

FIGURE 2.1. The cumulative number of satellite galaxies with $M_* > 10^5$ $M_\odot$ as a function of 3D distance around the MW (red) and M31 (black), similar to Fig. 2 from Yniguez et al. 2014. M31's line is lighter where the data are known to be incomplete at this stellar mass limit. Shaded regions are the 68 per cent and 95 per cent uncertainty in radial distribution when considering the line-of-sight distance uncertainties for satellites. Typical 68 per cent (95 per cent) scatter for the MW is ±0.3 (±0.5) satellites while for M31 it is ±1.2 (±2.4) satellites. The profiles of the MW and M31 are strikingly similar within 150 kpc, but diverge beyond that, where completeness is uncertain. We do not attempt to correct the LG observations for completeness.

where available, and elsewhere use $M_*/L_V = 1.6$ (Martin et al., 2008; Bell & de Jong, 2001). We apply the same stellar mass limit and host-satellite distance limit to the MW and M31 as in our simulation satellite criteria ($M_* > 10^5$ $M_\odot$ and d ¡ 300 kpc). For the satellite galaxies around the MW we take sky coordinates and distances with uncertainties from McConnachie (2012). To model the effects of uncertainties in observed distances, we sample MW satellite distances 1000 times assuming Gaussian distributions for the uncertainties to generate a median radial profile with scatter around the MW (Figure 2.1).

We exclude the Sagittarius dwarf spheroidal galaxy from our MW sample, because it is undergoing significant tidal interactions and we do not believe our halo finder would correctly

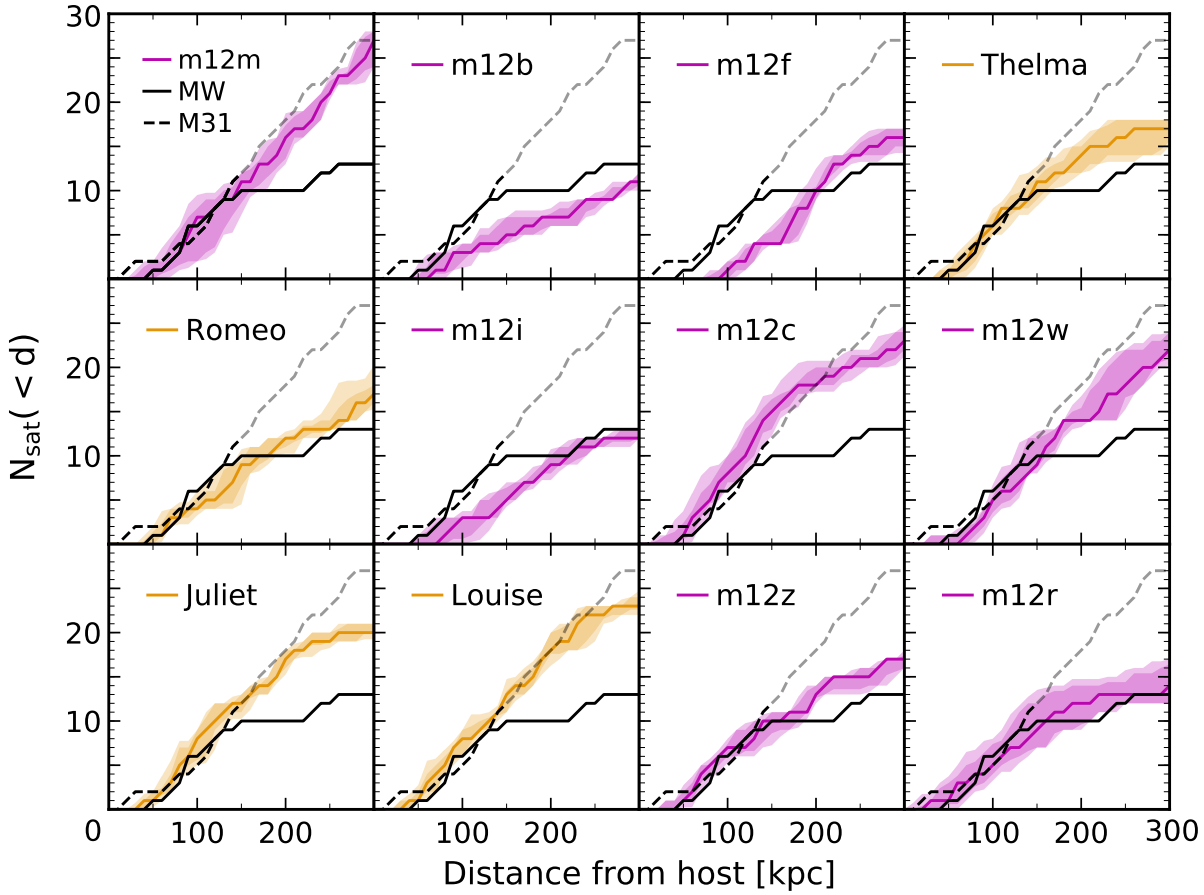

FIGURE 2.2. The cumulative number of satellite galaxies as a function of 3D distance from each host. Results are shown for satellite galaxies with $M_* > 10^5$ $M_\odot$. Colored lines are the median radial profile of the last 1.3 Gyr ($z = 0-0.1$, 11 snapshots in total), and the shaded regions are the 68 per cent and 95 per cent confidence intervals in variation over time. Isolated MW-like hosts are pink, while paired LG-like systems are orange. Black lines are the median profiles around the MW (solid) and M31 (dashed, lighter where incomplete), taking into account uncertainties in line-of-sight distance to satellites (see Figure 2.1 for scatter in observed profiles). The panels are ordered by decreasing stellar mass of each simulated host galaxy; m12m has the highest $M_* \approx 10^{11}$ $M_\odot$ while m12r the has lowest $M_* \approx 1.5 \times 10^{10}$ $M_\odot$. We do not see any obvious trend in simulated profile shapes or total number of satellite within 300 kpc as a function of host stellar mass. Across our sample we find simulated profiles that agree well with both the MW and M31.

identify it as a subhalo in the simulations. We also include two more recently discovered ultra-diffuse satellite dwarf galaxies of the MW: Crater 2 ($D_\odot \sim 118$ kpc; Torrealba et al. 2016) and Antlia 2 ($D_\odot \sim 130$ kpc; Torrealba et al. 2018), bringing the total number of MW satellites considered to 13. Using the nominal mass-to-light ratio of 1.6 we estimate the stellar masses of these additional galaxies to be $M_* \sim 2.6 \times 10^5$ $M_\odot$ for Crater 2 and $M_* \sim 3.4 \times 10^5$ $M_\odot$ for Antlia 2.

For the satellite galaxies around M31, we use sky coordinates where available from McConnachie (2012) and apply the same stellar mass and distance restrictions, leaving us with a total of 28 satellite galaxies. To obtain the 3D radial profiles of M31's satellites with uncertainties, we sample 1000 line-of-sight distances from the posterior distributions published in Conn et al. (2012), where available. However, several M31 satellites do not have published distance distributions: M32, NGC205, IC10, And VI, And VII, And XXIX, LGS 3, And XXXI, and And XXXII. In the cases of M32 and NGC205, they are too close to M31 to reliably determine their distances, so we assume they have the same line-of-sight distance distribution as M31 itself. Positions on the sky, distances, and distance uncertainties for And XXXI and And XXXII are taken from their discovery paper (Martin et al., 2013a). For the remaining satellites without line-of-sight distance posteriors, we sample the distances published in McConnachie (2012), assuming Gaussian distributions on the uncertainties.

Figure 2.1 shows the cumulative number of satellite galaxies around the MW and M31 as a function of 3D distance from the host, and the shaded regions represent estimated scatter in these profiles when we consider uncertainties in line-of-sight distance. While the sample for M31 includes 28 total satellite galaxies, when we include uncertainties the median number of satellites within 300 kpc is 27. The resulting 68 per cent scatter averaged across distance from host in LG radial profiles is ± 0.3 satellites for the MW and ± 1.2 satellites on average for M31. We discuss comparisons to the scatter in simulation profiles in Section 2.5.2.

Comparisons to the LG must also be understood in terms of observational completeness. However, the observational data used for comparison to the simulations in this work have *not* been completeness-corrected. From the Pan-Andromeda Archaeological Survey (PAndAS, McConnachie et al., 2009), the satellite population around M31 is complete to within 150 kpc (projected) of M31 down to half-light luminosities $L_{1/2} > 10^5$ $L_\odot$ (Tollerud et al., 2012). This includes our lowest satellite galaxy stellar mass limit (10^5 $M_\odot$), so we think we are making a fair comparison to M31 at least within 150 kpc (where we find evidence for tidal disruption of galaxies by the host, see Sections 2.5.5 and 2.5.6). However, if we assume that our simulations are representative of the LG we find that there may be more galaxies to discover around M31 beyond 150 kpc (see Section 2.5.9). Given that M31 already has a somewhat high number of satellite galaxies compared to the MW, this could potentially make M31's satellite population larger than those of the simulations used here.

Completeness around the MW is complicated by varied survey coverage and seeing through the Galactic disk (Kim et al., 2018). However, these sources of incompleteness are likely to affect only satellite galaxies fainter than classical dwarf galaxies and therefore they are unlikely to significantly change the results of this work. Of some concern is the proper identification of diffuse or low surface brightness galaxies (especially through the disk), but this is already being addressed using Gaia data to identify dynamically coherent stellar structures (like the Antlia 2 galaxy included in this work). We cannot preclude the possibility of further observational incompleteness down to our lowest stellar mass cut out to 300 kpc around the MW and M31. For this reason, we present comparisons at multiple (higher) stellar mass limits (Sections 2.5.1 and 2.5.4) and make predictions for the numbers of satellites to potentially be discovered around the MW and M31 (Sections 2.5.8 and 2.5.9).

We also compare our simulations and observations of the LG to results from the Satellites Around Galactic Analogs (SAGA) survey (Geha et al., 2017). SAGA targets MW analogs down to the luminosity of the Leo I dwarf galaxy ($M_r < -12.3$ or $M_* \approx 5 \times 10^6$ $M_\odot$), and the

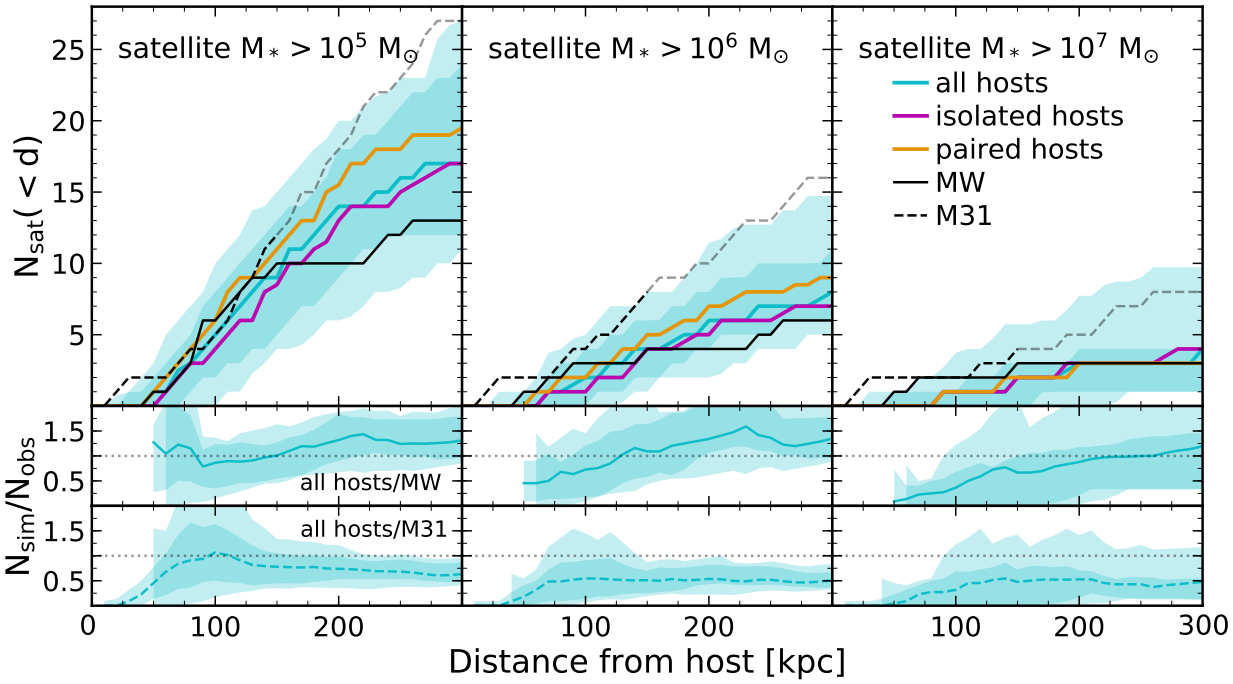

FIGURE 2.3. *Top row*: The cumulative number of satellite galaxies across all the simulations and snapshots as a function of 3D distance from the host, for satellites with $M_* > 10^5$ $M_\odot$ (left), $M_* > 10^6$ $M_\odot$ (middle), and $M_* > 10^7$ $M_\odot$ (right). Solid colored lines are the simulation median radial profiles over the last 1.3 Gyr ($z = 0 - 0.1$, using 11 snapshots), while the shaded regions show the 68 per cent and 95 per cent variation. We consider all simulations (blue), only the isolated hosts (pink), and only the LG-like paired hosts (orange). Black lines are the median radial profiles around the MW and M31, taking into account uncertainties in line-of-sight distance measurements. For the two lowest mass bins, the paired hosts have slightly more satellites on average, though this is within the 68 per cent scatter. The variation in simulation profiles spans the profiles of the MW and M31 for all three satellite stellar mass bins. *Bottom rows*: The median and scatter for all hosts' radial profiles normalized to the observational data for the MW (middle) and M31 (bottom). The simulation-to-MW ratio agrees with unity within the 68 per cent scatter at nearly all distances and for all satellite stellar mass limits. The simulation-to-M31 ratio agrees with or is close to unity within the 95 per cent scatter at most distances ($\gtrsim$50 kpc) for all satellite stellar mass limits.

initial results include the 2D radial profiles of satellite galaxies around 8 MW analogs within 20-40 Mpc of the LG. For more details on how we made this comparison, see Section 2.5.3.

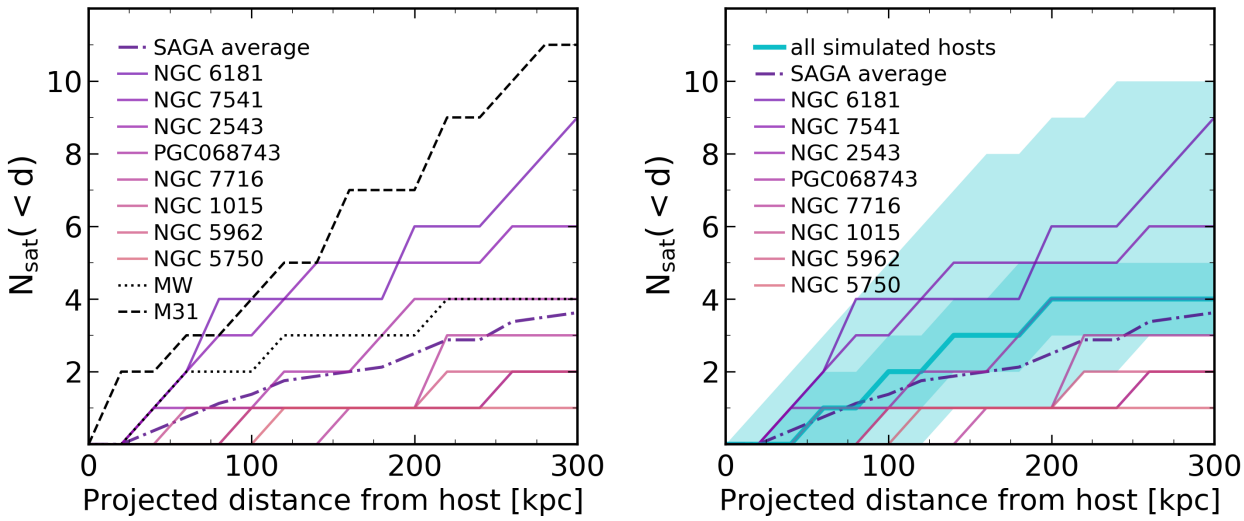

FIGURE 2.4. *Left:* Cumulative number of satellite galaxies with $M_* > 5 \times 10^6$ $M_\odot$ as a function of 2D projected distance, for observations of the LG (black lines) and the 8 complete MW analogs from the SAGA survey and their average (colored lines). The MW lies in the middle of the range of observed profiles, 1-2 satellites above the SAGA average, while M31 is at the upper edge of the distribution of profiles. *Right:* Same as left, but showing only SAGA profiles and 2D simulation median profile with scatter (blue). The scatter in the simulations is from random lines of sight, host-to-host variation, and variation over time (but time scatter is not significant). Three SAGA hosts have fewer satellites than the 95 per cent simulation limits, but the SAGA average lies mostly within the 68 per cent simulation scatter (and is always within the 95 per cent simulation scatter).

2.5. Results

We analyze satellite galaxy positions in our simulations over time using halo catalogs from 11 snapshots, taken over $z = 0 - 0.1$ ($\sim$1.3 Gyr) in steps of $\Delta z = 0.01$. We do this for each of the 12 simulated hosts, providing a total of 132 radial distributions of satellite galaxies at different times to study. In the inner halo, a typical satellite can undergo a full orbit in under 1 Gyr, while it may take $\sim$3-4 Gyr for a complete orbit in the outer halo. This time baseline allows us to time-average over satellite orbits to minimize sampling noise over at least 1/4 of an orbit, which is especially important at small distances where satellites spend the least amount of time. Our choice is motivated by a compromise between sampling sufficiently

across orbital histories and avoiding systematic redshift evolution (compared with $z = 0$) in the satellite populations. We find that time-averaging is critical for obtaining accurate results (see Section 2.5.2 for results on time variation in radial profiles).

2.5.1. Radial profiles. Figure 2.2 shows the cumulative number of satellite galaxies with $M_* > 10^5$ $M_\odot$ as a function of 3D distance from the host, or the radial profile, for each individual host-satellite system. The solid, colored lines are the simulated median radial profile across $z = 0 - 0.1$, and the shaded regions show the 68 per cent and 95 per cent variation over time. The median number of satellites within 300 kpc for the simulated hosts ranges from 11-27, consistent with the observed total number of $M_* > 10^5$ $M_\odot$ satellite galaxies within 300 kpc of the MW (median $N_{sat} = 13$) and M31 (median $N_{sat} = 27$) today. Hosts are ordered by stellar mass with m12m being the most massive ($M_* \approx 10^{11}$ $M_\odot$) and m12r the least massive ($M_* \approx 1.5 \times 10^{10}$ $M_\odot$). The number of satellite galaxies does not have an obvious correlation with host mass. The hosts show a wide range of profile shapes: m12m, m12c, m12w, and Louise closely follow M31, while Thelma, Romeo, m12i, m12z, and m12r more closely follow the MW.

Figure 2.3 summarizes the key result of this work: the radial profiles of satellite galaxies around the 12 hosts in our simulations span the observed radial distributions of satellites in the LG. Figure 2.3 aggregates all of our simulated profiles at three different satellite stellar mass thresholds: $M_* > 10^5$ $M_\odot$ (left), $M_* > 10^6$ $M_\odot$ (middle), and $M_* > 10^7$ $M_\odot$ (right). In the top panels we show the median and scatter across all 132 radial profiles simultaneously. The median radial profile for all simulated hosts (blue) lies on top of the median LG observations at distances < 150 kpc (where observational completeness is more secure), and at larger distances it lies between the MW and M31 profiles. The median for paired hosts (orange) is slightly above the total median (blue), while the median for isolated hosts (pink) is slightly below the total median. However, the paired and isolated medians are still within the total 68 per cent scatter across all the simulations.

The 68 per cent scatter in the simulations overlaps the 68 per cent scatter in MW observations (shown in Figure 2.1) at nearly all distances, and the MW's median profile is always within the 95 per cent simulation scatter. M31's median profile lies within the 95 per cent simulation scatter at nearly all distances. However, M31 appears to have a slight excess of satellites compared to the 68 per cent simulation scatter at small distances (<50 kpc) and large distances (>250 kpc) for all satellite M_* thresholds, though the uncertainties in M31's profile at small distances are relatively high (see Section 2.4 for more details). The 95 per cent scatter in simulations overlaps with the 68 per cent scatter in LG profiles at all distances (not shown here, but see Figure 2.1). We also a differentially-binned radial distribution for satellites with $M_* > 10^5$ $M_\odot$ in Appendix A.2, where we also see general agreement the LG and our simulations. We conclude that our simulation sample broadly agrees with and spans the profiles around the MW and M31.

In the bottom panels of Figure 2.3, we normalize the total simulation median and scatter to the MW and M31 radial profiles. To calculate the simulated-to-observed ratios, we divide the time-averaged radial profile of each of the 12 simulated hosts by 1000 sampled observational radial profiles of the MW or M31. Thus, the scatter in each of the bottom panels is from simulated host-to-host variation as well as observational uncertainties. We find that the MW ratio is consistent with unity at the 68 per cent level at nearly all distances for satellites with $M_* > 10^5$ $M_\odot$ and $M_* > 10^6$ $M_\odot$. This consistency breaks down at distances <150 kpc for $M_* > 10^7$ $M_\odot$ given the presence of the LMC and SMC, which are currently near their pericentric passage around the MW.

The M31 ratio is consistent with unity at the 95 per cent level at most distances for satellite $M_* > 10^5$ $M_\odot$ and $M_* > 10^7$ $M_\odot$, while for $M_* > 10^6$ $M_\odot$ the upper scatter in the ratio typically reaches ∼0.8. The M31 ratio is consistent with unity at the 68 per cent level within 50-150 kpc of the host for satellite galaxies with $M_* > 10^5$ $M_\odot$. Beyond 50 kpc, the median M31 ratio is typically ∼50 per cent across the different mass thresholds, indicating

that it has a somewhat large satellite galaxy population compared to our average simulation. This excess of satellite galaxies around M31 relative to the simulations is consistent at all distances, suggesting that M31 may just have more satellites overall, which may mean that its host halo mass is higher than in our simulated sample. The M31 ratio is most consistent with unity for our lowest mass bin and within 50-100 kpc. We interpret this, along with our resolution tests in Appendix A.1, as evidence that we are resolving our sample well even at these lower satellite masses.

Finally, to statistically test whether our simulations' radial profiles are consistent with the LG, we perform a two sample Kolmogorov-Smirnov (KS) test between the median profiles of the LG and each simulated host's profiles (at all 11 snapshots) for satellite galaxies with $M_* > 10^5$ $M_\odot$. The KS test compares the overall shape of the radial profile, and is less sensitive to the absolute number of satellites than taking a ratio between simulations and observations. We calculate the KS statistic for each of the 11 snapshots over $z = 0 - 0.1$, and we quote the percentage of snapshots where a simulation was *inconsistent* with either the MW or M31. The KS test results show that a few of the simulations are *inconsistent* with being drawn from the same distribution as the MW at a significance level of 95 per cent: m12f (83 per cent), m12m (27 per cent), m12i (18 per cent), and m12w (9 per cent). Only m12r (9 per cent) is inconsistent with M31's distribution, and only at one of the 11 snapshots. We also use the Anderson-Darling (AD) test to check these results and maximize sensitivity to the tails of the radial distributions. With the AD tests, we achieve essentially the same results as the KS tests. We also repeat the KS and AD tests for satellite galaxies with $M_* > 10^6$ $M_\odot$, and found that none of the simulated profiles are inconsistent with the MW or M31 at the 95 per cent level, possibly indicating even better agreement at higher masses and that simulations and observations are well resolved and complete in this mass range.

2.5.2. Scatter across hosts versus across time. Table 2.1 summarizes host galaxy mass, number of satellites per host within representative distances, and the scatter over time in each host's radial profile. We quantify the scatter in radial profiles using the 68 per cent scatter around the median number of satellites with $M_* > 10^5$ $M_\odot$ within a given distance from their host. To understand the importance of time versus host-to-host scatter, we compare the radial profile scatter within individual hosts over time (the pink and orange shaded regions from Figure 2.2), scatter among hosts after their time dependence has been averaged out (the solid, colored median lines in Figure 2.2), and total scatter among all hosts and snapshots simultaneously (the blue shaded region of Figure 2.3). We quote the 68 per cent scatter about the median in absolute number of satellites and also quote scatter as a percentage relative to the median to give an idea of the fractional variation. We consider the scatter at three different distances (50, 100, and 300 kpc) to measure time dependence over the full range of the radial profiles.

First, we consider the scatter in the total number of satellite galaxies within 300 kpc. The combined scatter across all hosts and snapshots within 300 kpc is $\pm$6 satellites, or a 35 per cent variation when normalized to the median of 17 satellites. The host-to-host scatter after averaging time dependence out is $\pm$4.7 satellites (30 per cent), whereas the average scatter over time for an individual host is much lower at $\pm$1.1 satellites (5 per cent). Thus we find that total scatter at large distances is dominated by to host-to-host variations rather than time dependence.

Within 100 kpc, the combined scatter across hosts and time is $\pm$3 satellites (60 per cent), while the host-to-host scatter is $\pm$2.5 satellites (50 per cent), and the time scatter is $\pm$1.3 satellites (25 per cent). The increased fractional significance of time scatter is likely caused by the relatively small amount of time that satellites spend near pericenter of their orbits. Within 50 kpc we see that this effect is exacerbated: the combined scatter across hosts and time is $\pm$1 satellite (100 per cent), while host-to-host scatter is $\pm$0.5 satellites (50 per cent),

and time scatter is ±0.6 satellites (60 per cent). We conclude that at large distances ($\gtrsim$100 kpc) the total scatter across all 132 radial profiles is dominated by host-to-host variation, and at small distances ($\lesssim$50 kpc) the total scatter is dominated by time dependence from satellite orbits.

2.5.3. Comparison to the SAGA survey. We also compare our simulated and LG profiles to the on-the-sky projected radial profiles of 8 MW analogs in the SAGA survey. To match the SAGA luminosity limit, we select satellite galaxies for comparison to SAGA in our simulations, the MW, and M31 by requiring them to have stellar masses above the value of Leo I, $M_* \approx 5 \times 10^6$ $M_\odot$. We generate 2D projections of the simulations along 1000 lines of sight for each of the 12 host-satellite systems at 11 snapshots, from which we compute the median and scatter across the simulated sample. For M31 satellites, we use only their projected on-the-sky distances from M31, assuming a line-of-sight distance to M31 of 780 kpc. For the MW, we use the 3D positions of the satellites and their line-of-sight distance uncertainties to generate 2D realizations from 1000 lines of sight as we did for the simulations.

Figure 2.4 (left) shows the observed 2D profiles for SAGA hosts, the MW, and M31. Most SAGA systems have fewer satellite galaxies compared to the MW and M31, which could be an effect of the broad mass selection function used in SAGA to choose MW analogs within uncertainties on the MW's stellar mass (Geha et al., 2017). Because our simulations show only slightly higher satellite counts in our LG pairs compared with isolated hosts (see Figure 2.3), this implies that the SAGA selection of *isolated* hosts is unlikely to be a significant cause of difference as compared with the LG. The MW profile lies in the middle of the SAGA sample, and its scatter via line-of-sight averaging spans most of the range between the average SAGA profile and M31's profile within 200 kpc. M31 still has a relatively large number of satellites compared to the SAGA observations at all distances (especially beyond 150 kpc), but two of the SAGA hosts have numbers of satellites approaching M31's profile.

Figure 2.4 (right) shows the SAGA profiles compared to the simulations. The blue line is the median and the shaded regions show the 68 per cent and 95 per cent scatter in simulations. The scatter in simulated 2D profiles is mainly due to host-to-host variation and line-of-sight averaging, while time variation contributes a negligible amount of scatter in projection. At distances >100 kpc, three of the eight SAGA hosts are at or below the 95 per cent simulation limits. The SAGA average lies within the 68 per cent simulation scatter at most distances, though still slightly below the simulation median for distances >100 kpc. The best agreement between the SAGA average and the simulations is at small distances (<100 kpc), where they overlap the most. Overall, the simulation scatter encompasses five of the eight SAGA profiles and we find reasonable agreement among SAGA results, the LG, and our simulations in projection.

2.5.4. Dependence on satellite galaxy stellar mass. In this section, we examine in more detail whether our results within small distance ($\lesssim$100 kpc) depend on the stellar mass of satellite galaxies. This is a test of how our simulations compare to observations across our satellite mass range, and because higher-mass satellites are better resolved in both stellar mass and halo mass, it is also a test of dependence on resolution. Here, we assume that satellites with larger stellar masses inhabit more massive dark matter halos, but this may not always be true given scatter in the galaxy stellar mass-halo mass relation (Garrison-Kimmel et al., 2017a; Fattahi et al., 2018). We use the number of satellite galaxies within small distances as our summary statistic because this is where we expect to see the most prominent effects of tidal disruption and perhaps numerical over-merging in simulations. However, given the small numbers of satellites within 50 kpc of the hosts in both our simulations and the observations, we choose 100 kpc as the limiting distance as a reasonable trade-off between testing at small distances and obtaining reasonable statistics.

The top panel of Figure 2.5 shows the median number of simulated satellite galaxies within 100 kpc of their host across all hosts and snapshots as a function of the lower limit

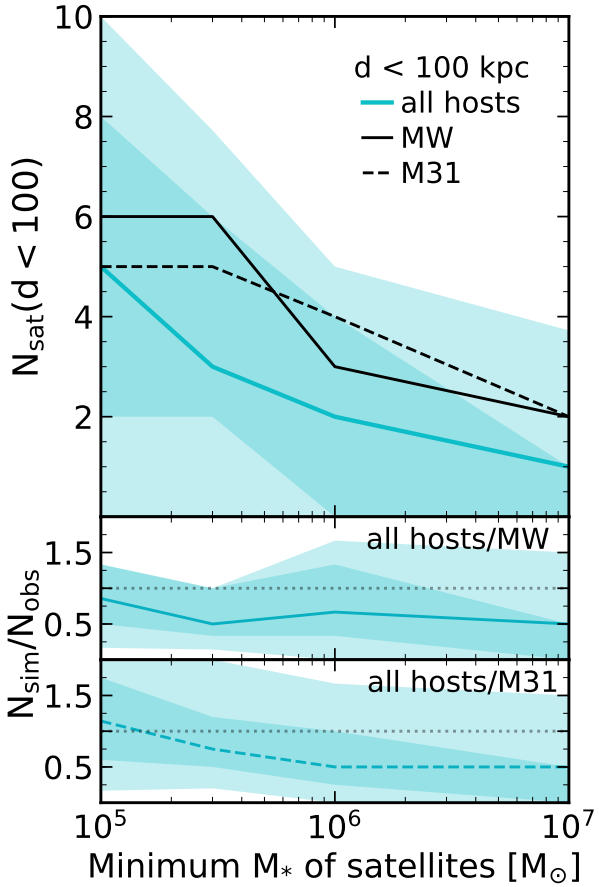

FIGURE 2.5. *Top panel:* The number of satellite galaxies with $M_* > 10^5$ $M_\odot$ within 100 kpc of their host as a function of minimum stellar mass. The blue line and shaded regions show the median, 68 per cent, and 95 per cent variation over the last 1.3 Gyr ($z = 0 - 0.1$, using 11 snapshots) across all of the simulations. The simulation median is $\lesssim$2$\times$ smaller than observations of the MW and M31, but the scatter in simulations encompasses the MW and M31 at all satellite masses. *Bottom panels:* The median and scatter in the ratio of N_{sat}(<d) in the simulations relative to observations of the MW and M31. The trend in the ratios is essentially flat with increasing minimum satellite stellar mass. Even if the simulations have fewer satellites on average within 100 kpc, less massive satellites (hence closer to the resolution limit) are not preferentially under-represented or over-disrupted in the simulations compared to observations.

on satellite stellar mass, compared to the MW and M31. We consider satellite stellar mass limits from $M_* > 10^5$ $M_\odot$ up to $M_* > 10^7$ $M_\odot$, the highest stellar mass bin where we still have sufficient statistics in our simulations. The observed medians for the MW and M31

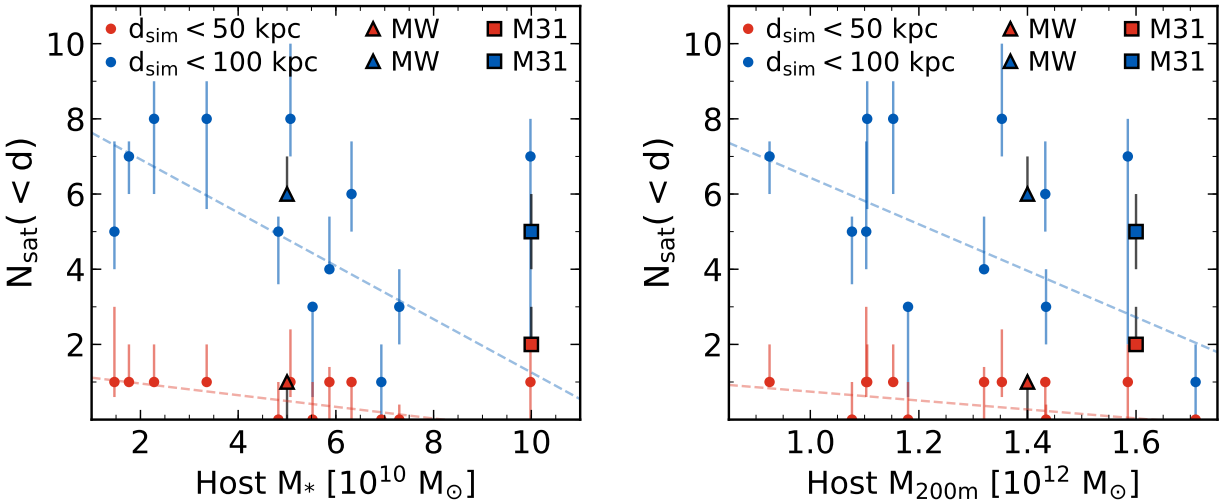

FIGURE 2.6. The number of satellite galaxies ($M_* > 10^5$ $M_\odot$) within 50 (red) and 100 (blue) kpc of each host as a function of host stellar mass (left) and host halo mass (right). Circles with error bars are the simulated host medians and 68 per cent variation over the last 1.3 Gyr ($z = 0 - 0.1$, using 11 snapshots). Observations of the MW are triangles and observations of M31 are squares, and their error bars are from uncertainties in line-of-sight distances. We use a linear fit to the simulations to demonstrate the negative trends. *Left:* The number satellites decreases with increasing host stellar mass within both 50 and 100 kpc of the host. The red points show that there are little to no satellites within 50 kpc in the simulations. The MW has a number of satellites comparable to the simulations, and M31 is within simulation variation at the high mass end. *Right:* Same as left, but using the halo mass (M_{200m}) of the host. There are similar trends in the number of satellite galaxies nearby their host, but the number of satellite galaxies within 50 kpc is less correlated with halo mass than it is with stellar mass. Though not shown, the number of satellite galaxies within 300 kpc is essentially uncorrelated with host stellar and halo mass (see Fig. 3 of Garrison-Kimmel et al. 2018 for satellite counts within 300 kpc as a function of host halo virial mass). This indicates that the host's stellar mass is a better predictor of the survival of satellite galaxies within 50 kpc.

are $\lesssim$2$\times$ higher than the simulation median. This difference for satellites with $M_* > 10^7$ $M_\odot$ could be caused by the presence of the LMC and SMC near their pericenters around the MW, which is not typical in a time-averaged sense. Even so, the 95 per cent simulation

scatter always encompasses the observations, and the 68 per cent scatter mostly contains the MW and M31 lines.

In the bottom panels, we normalize our simulations to the MW and M31 observations, by sampling from both the simulated hosts and observational uncertainties simultaneously. In general, satellite galaxies with smaller stellar masses reside in less massive dark matter halos, so they are resolved with fewer star and dark matter particles. Therefore, in the absence of confounding numerical artifacts, we might expect our simulations to show increasingly fewer satellites relative to observations at lower stellar masses if we are reaching our resolution limit. Interestingly, we find the best agreement with observations when we include our lowest mass satellite galaxies ($M_* > 10^5$ $M_\odot$). The simulated-to-observed ratios are always consistent with unity at the 95 per cent level, but are only consistent with unity at the 68 per cent level when we include satellites with $M_* > 10^{5-6}$ $M_\odot$. The trend in the ratios as a function of minimum satellite stellar mass considered is relatively flat, though our simulations may not be producing as many higher-mass satellites as the LG. This is broadly consistent with results from Garrison-Kimmel et al. (2018), who examined all satellites out to 300 kpc and found that most hosts are consistent with the MW and M31's satellite population is only slightly larger than the simulations. Therefore, relative to observations, our simulations do not suffer from obvious over-destruction of satellites at the stellar masses that we consider.

To more thoroughly analyze numerical resolution, we test for convergence of the radial distributions of subhalos in these simulations compared to those from lower resolution simulations in Appendix A.1. There, we show that our (high resolution) simulations are converged to within $\sim$ 20 per cent on average, and the 68 per cent (host-to-host) scatter is consistent with 100 per cent convergence at distances $> 30 - 40$ kpc. We note that this exercise suffers from the effects of an additional disruptive effect in the low-resolution simulations: because the low-resolution host galaxies have $\sim 2\times$ larger stellar masses, their subhalos may be more easily tidally stripped or destroyed as they orbit close to the host. This may

have the effect of making the high-resolution simulations appear less converged, at least at small distances from the host. We conclude that subhalos hosting the satellite galaxies in our high-resolution simulations are sufficiently resolved for tests of the satellite populations' radial distributions. For a more nuanced discussion of convergence and additional resolution tests, see Appendix A.1.

2.5.5. Dependence on host mass. We test whether our results for satellite galaxies with $M_* > 10^5$ $M_\odot$ are sensitive to the stellar and halo masses of the host galaxies within 50 and 100 kpc. Figure 2.6 (left) shows the median number of satellite galaxies within 50 and 100 kpc of the host as a function of host stellar mass. The simulations agree well with the MW, and while the simulation trends lie below M31, the scatter for the most massive simulated host (m12m) is still consistent with M31. Within both 50 and 100 kpc of the host there is a negative trend in the number of satellites as a function of host galaxy stellar mass, and 4 hosts have no satellites at all (median over time) within 50 kpc. These 4 hosts all have stellar masses $\gtrsim 5 \times 10^{10}$ $M_\odot$, which is the average host stellar mass for the simulations. We interpret this and the trend lines as evidence for enhanced tidal destruction of satellites in our simulations due to the increased gravitational potential from the more massive hosts' baryonic disks.

Figure 2.6 (right) shows the median number of satellites within 50 and 100 kpc as a function of host halo mass (M_{200m}). When controlling for host halo mass, the time variation or scatter in the simulations is consistent with the MW and M31. However, M31 lies above both the simulation trends and the MW lies slightly above the 100 kpc trend line. M31 on the other hand, lies above the simulation trends, but still within the simulation scatter. The trend in N_{sat}(d¡100 kpc) as a function of host halo mass is slightly less steep than as a function of host stellar mass for the simulations. The correlation of N_{sat}(d¡50 kpc) with host mass is stronger for stellar mass (Pearson correlation coefficient: $r_* = -0.32$) than it is for halo mass ($r_{200m} = -0.22$). The correlations of N_{sat}(d¡100 kpc) with each type of host

mass are: $r_* = -0.35$ and $r_{200\mathrm{m}} = -0.43$. Within 300 kpc (not shown) we find little to no correlation: $r_* = 0.14$ and $r_{200\mathrm{m}} = 0.07$. We interpret the larger correlation with host stellar mass within 50 kpc and steeper trend with host stellar mass within 100 kpc as the destructive tidal effects of the host baryonic disk manifesting at sufficiently small distances. Since host stellar mass is more correlated with satellite count within 50 kpc, we conclude that host stellar mass is a better predictor of the total number of surviving satellite galaxies within 50 kpc of the host, where we expect disk effects to be strongest.

Naively, we might expect the number of satellites at any distance to correlate positively with halo mass, and because M_* correlates with $M_{200\mathrm{m}}$, we might also expect a similar correlation with stellar mass. Both the negative trend with host stellar mass and the lack of satellites around the more massive galactic disks suggest instead that the baryonic disk is depleting the satellite population at small distances. However, we note that because of the correlation between host M_* and $M_{200\mathrm{m}}$ in our simulations (see Figure A.3), we cannot strictly disentangle the tidal effects of the separate disk and halo components of the host independently in our analysis. Despite this uncertainty, we find it physically plausible that tidal destruction of satellites can negate our initial expectations, at least for satellites closer to the host galaxy, consistent with results presented in Garrison-Kimmel et al. (2017b) and Kelley et al. (2019) that show a lack of satellites or subhalos at small distances in the presence of a disk potential. This also explains the lack of correlation between the number of satellites within 300 kpc and host halo mass (also noted in Fig 3 of Garrison-Kimmel et al. 2018 as a function of host halo virial mass): while increasing halo mass increases the number of expected satellites, the correlation of host stellar mass with host halo mass and the tidal destruction from the host disk act to cancel out this dependence, at least within the limited host mass range that we explore with our simulations.

We also note that while our simulated hosts have a wide range of stellar masses ($M_* \sim 10^{10-11}$ $M_\odot$), they were selected over only a narrow range in host halo mass ($M_{200\mathrm{m}} \sim$

$1 - 2 \times 10^{12}$ $M_\odot$). Therefore, our sample is missing MW/M31-like host galaxies with much larger (or smaller) halo masses, but with stellar masses that scatter into our sample's range. Hosts with more extreme halo masses like this could potentially lead to a less negative correlation of N_{sat} with host M_* within 100 kpc.

2.5.6. Comparison with dark matter-only simulations. The seven Latte simulations also have dark matter-only (DMO) versions run with the same number of DM particles and the same force softening[2]. We compare the DMO versions to the baryonic simulations in order to investigate the effects of baryonic physics on the radial profiles of subhalos. To compare with the baryonic simulations, we find that satellite galaxies with $M_* > 10^5$ $M_\odot$ have typical peak dark matter halo masses $M_{peak} \gtrsim 8 \times 10^8$ $M_\odot$.

We select subhalos in the DMO and baryonic simulations by requiring them to be within 1000 kpc of their host and to have $M_{peak} > 8 \times 10^8$ $M_\odot$ so their halos are approximately as well-resolved as baryonic satellites down to $M_* \sim 10^5$ $M_\odot$. We then average the radial profiles of each host-subhalo system over $z = 0 - 0.1$ using all available snapshots (67 total) for improved subhalo statistics at small distances. We compute the ratio of a host's baryonic-to-DMO profiles for each host individually, and then examine the median and scatter across hosts. We compute the ratio as a function of distance for both cumulative and differential subhalo counts: $N(< d)$ and $N(d_1 < d <= d_2)$, respectively. Figure 2.7 shows the cumulative ratio of baryonic-to-DMO subhalos (top) and the differential ratio (bottom). The line and shaded regions are the median and scatter showing only host-to-host variation, as the time dependence has been averaged out prior to taking the ratio.

Within ∼100 kpc from the hosts, there are $\lesssim$50 per cent the number of baryonic subhalos compared to DMO subhalos, and this continues to rapidly drop as distance decreases until the (median) ratio reaches zero at 10-15 kpc. As Garrison-Kimmel et al. (2017b) studied

[2]However, the DMO simulations have DM particles with slightly higher masses of $m_{dm} = 4.2 \times 10^4$ $M_\odot$ due to the lack of baryons. We correct for this by multiplying DMO subhalo masses by $1 - f_b$ to account for the mass that would be otherwise relegated to baryons given the cosmic baryon fraction ($f_b \equiv \Omega_b/\Omega_m$) of our baryonic simulations.

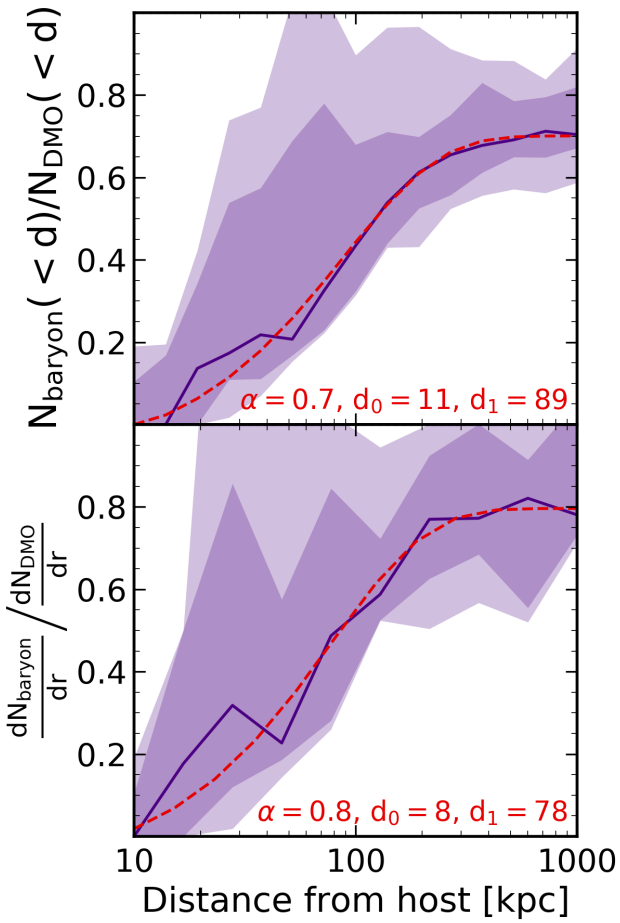

FIGURE 2.7. The ratio of the number of subhalos, at a given subhalo mass, in baryonic versus dark matter-only (DMO) simulations. Purple line and shaded regions are the host-to-host median and scatter in the baryonic-to-DMO ratio as a function of 3D distance. Red line is an analytic fit to the ratio, and fit parameters are also shown in red. Subhalos were selected to have $M_{peak} > 8 \times 10^8$ $M_\odot$, to mimic the halo masses of dwarf galaxies in the baryonic runs. *Top panel:* Baryonic-to-DMO ratio for cumulative subhalo counts as a function of distance. Relative to the DMO simulations, the baryonic simulations have ∼70 per cent (median) the number of subhalos beyond 200 kpc. The ratio drops rapidly within this distance, where the DMO subhalos are not subject to the additional gravitational potential of a host's baryonic disk. The median ratio declines to zero within ∼15 kpc of the host. *Bottom panel:* Same as above, but showing differential subhalo counts (discrete distance bins). The ratio is ∼80 per cent (median) beyond 200 kpc, and it declines to zero within ∼10 kpc of the host.

extensively using embedded disk potentials in DMO simulations of m12i and m12f, this is almost entirely due to the presence of the additional gravitational potential from the disk in the baryonic simulations. Here, we provide a more robust sample of simulations where we also time-average for each host, which is critical given how little time satellites spend near pericenter. The scatter within 100 kpc is greater than the scatter at 200-1000 kpc, due to a few hosts that have a number of baryonic subhalos closer to their number of DMO subhalos at small distances.

At large distances ($\gtrsim$200 kpc), the median ratios of baryonic-to-DMO subhalos flatten to $\sim$0.7 for the cumulative case and $\sim$0.8 for the differential case. This indicates that baryonic effects can reduce the masses of halos even at large distances from the host by $\sim$20-30 per cent as compared with DMO simulations. The overall reduction of substructure in the baryonic simulations relative to the DMO simulations is likely due to a combination of various baryonic effects such as reionization through our UV background and environmental effects like ram-pressure stripping and interactions with large scale structure such as filaments (Benítez-Llambay et al., 2013). Any of these processes may act to blow out gas from the galaxies in our baryonic simulations, shallowing their gravitational potential significantly in lower-mass galaxies like dwarfs, which in turn allows for easier removal of dark matter mass through gravitational interactions. Sawala et al. (2017) also found that the abundance of for subhalos with masses below $10^{9.5}$ $M_\odot$ in the `APOSTLE` simulations was reduced at all distances out to 200 kpc from the hosts. For the largest distances they examine, between 50 and 200 kpc, Sawala et al. (2017) found a reduction in substructure abundance of 23 per cent. This is similar to our results for the ratio of baryonic-to-DMO differential profiles between about 200 to 1000 kpc, where where we see a reduction in substructure abundance of about 20 per cent.

We provide fits to the ratio of baryonic-to-DMO subhalo counts as a function of distance that may be used to estimate the number of subhalos containing satellite galaxies ($M_* > 10^5$

$M_\odot$) in other DMO simulations. We fit to the median ratio across hosts, and use the 68 per cent variation in the ratio as uncertainty on the fitted median values[3]. In Table 2.2, we also explore fits using other subhalo mass cuts. We fit the median of the cumulative and differential baryonic-to-DMO ratios as a function of distance (d):

$$f(d) = \begin{cases} 0 & 0 \leq d < d_0 \\ \alpha \left[1 - e^{-\frac{d-d_0}{d_1}}\right] & d \geq d_0 \end{cases} \tag{2.1}$$

Where α is the asymptotic value of the ratio for infinitely large d, d_0 is the inner cutoff where the ratio goes to zero, and d_1 is the distance within which the ratio sharply declines. For the cumulative profile shown we find: $\alpha = 0.7$, $d_0 = 11$ kpc, and $d_1 = 89$ kpc. For the differential profile shown we find: $\alpha = 0.8$, $d_0 = 8$ kpc, and $d_1 = 78$ kpc. Table 2.2 shows these parameters for other fits using instantaneous bound halo mass for subhalo selection (not shown in Figure 2.7).

We find that, as expected, the fitted baryonic-to-DMO subhalo count ratios tend towards close to unity at large distances and drop to zero near the baryonic disk boundary. The fits indicate that even at arbitrarily large distances from the host, the baryonic subhalos are subject to additional destructive baryonic effects. The decline in the fitted ratios within $\sim$100 kpc is strikingly sharp: the cumulative and differential ratios both go to zero within ~ 10 kpc, indicating the physical boundary of intense gravitational effects from the baryonic disk. We see the same general trends in fits, for both cumulative and differential ratios, across the three different subhalo selection methods we use.

Our results agree with studies that have found that satellite survival depends on host-satellite distance at pericentric passage (e.g. D'Onghia et al., 2010; Zhu et al., 2016; Sawala

[3]The $z = 0$ snapshots of baryonic m12i, m12f, and m12m are publicly available at `ananke.hub.yt` for comparison to individual hosts.

et al., 2017; Garrison-Kimmel et al., 2017b; Nadler et al., 2018; Rodriguez Wimberly et al., 2019). We note that the destruction that we see is somewhat less extreme than in Garrison-Kimmel et al. (2017b), who used two of our baryonic simulations (m12i and m12f) and found no surviving subhalos at $z = 0$ within $\sim$20 kpc of the host. Our results here are more robust given the larger host sample and that we time-average the profiles.

Kelley et al. (2019) examined the destructive effects of an analytical disk+bulge potential embedded in DMO simulations, where the analytical potential was allowed to realistically grow over time to match the MW's potential at $z = 0$. They found the ratio of subhalo counts that were subject to the embedded potential relative to subhalo counts that were not subject to the additional potential to be $\sim$1/3 within 50 kpc of their hosts. We find that our baryonic simulations are more efficient at destroying subhalos within 50 kpc, with a median ratio of baryonic-to-DMO subhalo counts of $\sim$1/5 at this distance. This could mean that additional baryonic effects, such as supernovae, in our simulations lead to enhanced modulation of the baryonic-to-DMO ratio. However, the simulations used in Kelley et al. (2019) were calibrated to the mass of the MW and may not capture the full effects of our wider mass range which encapsulate more massive M31-like galaxies as well.

Sawala et al. (2017) performed a similar comparison of the radial distributions of substructure in baryonic and DMO simulations, averaging over time and four hosts from the `APOSTLE` simulations. However, the baryonic disks of the hosts in their simulations are $\sim 2 \times 10^{10}\ M_{\odot}$, which is lower than the average disk masks of our hosts. Thus, based on our results from Section 2.5.5, we expect to see more substructure destroyed around our hosts than Sawala et al. (2017) found. They found a baryonic-to-DMO ratio of subhalo counts of $\sim 1/2$ at $d < 10$ kpc, and $\gtrsim 3/4$ at $d > 50$ kpc. By comparison, we see a much smaller median baryonic-to-DMO ratio of zero within 10 kpc of our hosts, but the host-to-host scatter reaches as high as $\sim 1/5$ at $d < 10$ kpc. At $d > 50$ kpc, the scatter in our ratio varies from $\sim$ 1/5–1 and at $d > 100$ kpc it is $\gtrsim 1/2 - 1$. Newton et al. (2018) repeated this exercise and

TABLE 2.2. Parameters for fits to the ratio of subhalos in baryonic versus dark matter-only simulations in Equation 1. Cumulative distributions refer to the total number of subhalos enclosed as a function of 3D distance while differential distributions refer to discrete bins in 3D distance.

Subhalo selection method	α	d_0 [kpc]	d_1 [kpc]
Cumulative distributions			
$M_{peak} > 8 \times 10^8$ $M_\odot$	0.7	11	89
$M_{bound} > 10^8$ $M_\odot$	0.8	13	106
$M_{bound} > 10^7$ $M_\odot$	0.8	2	98
Differential distributions			
$M_{peak} > 8 \times 10^8$ $M_\odot$	0.8	8	78
$M_{bound} > 10^8$ $M_\odot$	0.9	21	95
$M_{bound} > 10^7$ $M_\odot$	0.9	0	100

found subhalo depletion similar to Sawala et al. (2017): their baryonic-to-DMO ratio ranged from $\sim 1/2$ at small distances and rose to $\sim 4/5$ at large distances (R_{200}) from the host. Considering the differences in the stellar masses of the host disks between our simulations, the larger subhalo depletion we see at small distances compared to that from Sawala et al. (2017); Newton et al. (2018) is unsurprising, and we note that far from the host disk our results are more similar to each other.

2.5.7. Radial concentration. We further quantify satellite radial profiles using their shape, which we refer to as radial concentration. A profile with higher concentration generally has more of its satellites at small distances than at large distances from the host. We parameterize the concentration of our simulated and observed radial profiles using two metrics: R_{90}/R_{50}, the ratio of the distances enclosing 90 per cent and 50 per cent of the total number of satellite galaxies around a host, and R_{90}/R_{10} to be sensitive to variations in satellite counts at smaller distances.

We analyze the concentration of 3D profiles considering LG satellites, baryonic simulation satellites, and DMO simulation subhalos that are within 300 kpc of their host. We measure concentration of the baryonic profiles for satellite galaxies with $M_* > 10^5$ $M_\odot$ around each of the 12 baryonic hosts and in the LG. DMO subhalos were selected as in Section 2.5.6,

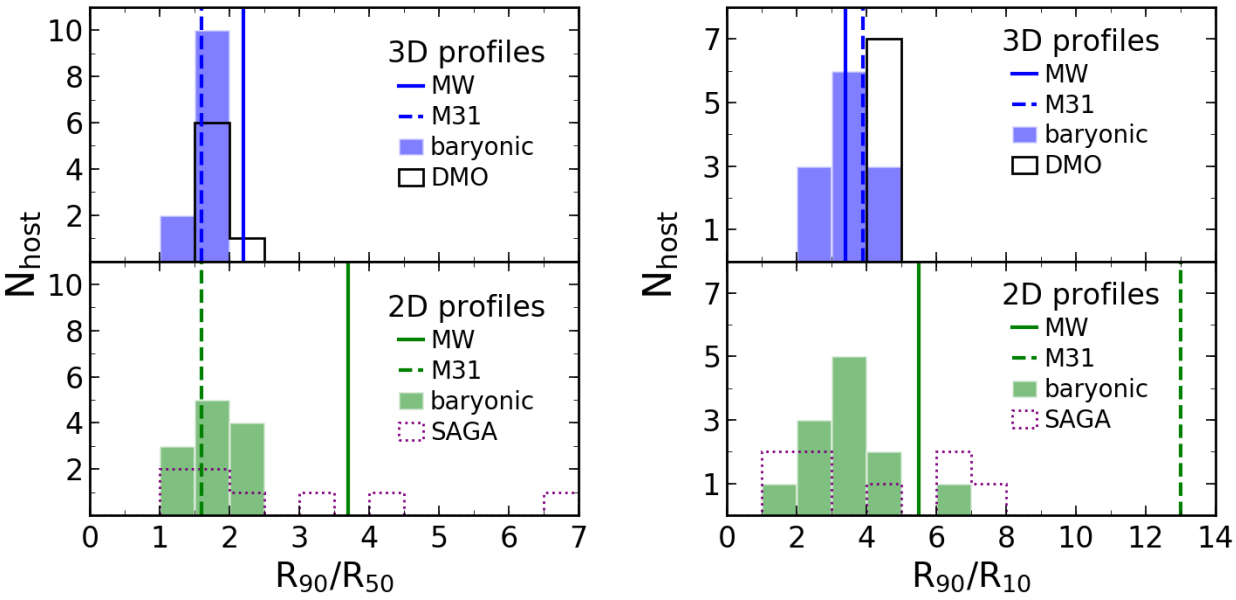

FIGURE 2.8. Radial concentration values of the simulated and observed profiles. Top panels are 3D profiles of all baryonic satellites with $M_* > 10^5$ $M_\odot$ ($M_{peak} > 8 \times 10^8$ $M_\odot$ for DMO), and bottom panels are 2D projections of profiles for satellites with $M_* > 5 \times 10^6$ $M_\odot$ for comparison to the SAGA survey. Filled color histograms are baryonic simulations, unfilled black histograms are DMO simulations, colored vertical lines are the MW and M31, and dashed unfilled histograms are SAGA systems. *Left:* Concentration as measured by R_{90}/R_{50}. The 3D simulated and observed profiles (top) have a narrow range of concentration values. The MW agrees better with the slightly higher concentrations of the DMO simulations, while M31 agrees with both the baryonic and DMO simulations. The 2D profiles (bottom) of the baryonic simulations, M31, and most of SAGA lie in the same narrow range as the 3D profiles. However, the MW and 3 of the SAGA hosts have much higher concentration. *Right:* Same as left, but for R_{90}/R_{10}. The 3D profiles are distributed over a narrow range in concentration with DMO simulations tending to have higher concentration. Both the MW and M31 agree with the baryonic simulations. The 2D profiles are spread over a wider range. The MW, 3 SAGA hosts, and one baryonic host have slightly higher concentration, while M31 is much more concentrated than any of the other systems in projection.

by requiring $M_{peak} > 8 \times 10^8$ $M_\odot$ for each of the 7 available DMO hosts. We also analyze the concentration of profiles in 2D projection for LG satellites, simulated baryonic satellites, and SAGA survey satellites with $M_* > 5 \times 10^6$ $M_\odot$. We report the concentration of each

simulated host as the median over 11 snapshots from $z = 0 - 0.1$, and the observed LG values as the median across 1000 sampled profiles.

Figure 2.8 (left) summarizes R_{90}/R_{50} concentration measurements for the baryonic simulations, DMO simulations, the LG, and the SAGA survey. R_{90}/R_{50} does not significantly differentiate baryonic simulations from DMO simulations. M31's R_{90}/R_{50} agrees with both the baryonic and DMO simulations, but the MW's R_{90}/R_{50} is slightly higher than the baryonic simulations, and is more consistent with the DMO simulations. However, we do find that $\sim$10-30 per cent of individual snapshots for half of the baryonic hosts (m12b, m12c, m12r, m12z, Romeo, and Thelma) have R_{90}/R_{50} values that are at least as concentrated as the MW. This suggests that the MW has a slightly more concentrated profile shape relative to the median values for each baryonic simulation host. In 2D projection, the MW appears more highly concentrated than the baryonic simulations and M31. Most of the 2D SAGA profiles over the baryonic simulation distribution, but two SAGA systems have much higher concentration that is closer to the MW and one SAGA system has a concentration nearly twice that of the MW.

Figure 2.8 (right) shows R_{90}/R_{10} concentration measurements for the simulations and observations. The baryonic simulations cover a broader range of values for R_{90}/R_{10} than they do for R_{90}/R_{50}. DMO simulations tend to have systematically higher average R_{90}/R_{10} than the baryonic simulations. Thus, the primary difference between baryonic and DMO profiles lies in the fraction of satellites at small distances ($\lesssim$100 kpc), where the DMO simulations have a larger fraction of their subhalos. This is consistent with the results of Section 2.5.6, where we show that the largest discrepancies between baryonic and DMO profiles occur within $\lesssim$100 kpc of the hosts. The R_{90}/R_{10} values for the MW and M31 are consistent with the baryonic simulations and lie outside the range of DMO values. The 2D profile span an even broader range in R_{90}/R_{10} than the 3D profiles. The SAGA systems are broadly consistent with the baryonic simulations, with a few more SAGA systems lying at

the high end of the baryonic distribution. The MW in projection is also near the higher end of the baryonic simulations, and M31 appears much more concentrated than anything else.

The concentrations of the MW and M31 profiles generally overlap with the concentrations of the simulated baryonic profiles. Considering incompleteness in M31's satellite population, if there are more M31 satellites to discover beyond 150 kpc, it could potentially push M31's R_{90} higher. This could increase M31's concentration to a point where it is discrepant with the baryonic simulations. However, when using the R_{90}/R_{50} metric, the MW is slightly more radially concentrated and therefore less consistent with the baryonic simulations than the DMO simulations. The MW in 2D projection appears more concentrated than most of the baryonic simulations, and M31 in projection is more concentrated than anything else. The SAGA profiles mostly overlap the projected baryonic simulation profiles, with a few SAGA systems having higher concentration more like the MW. Our results indicate that the MW may not be as much of a high-concentration outlier as previously thought: Yniguez et al. (2014) noted that the MW had a larger concentration than all of their DMO simulations. Concentration depends strongly on observational completeness assumptions though, and this may hint that there are more satellites just above $M_* = 10^5$ $M_\odot$ remaining to be discovered at farther distances from the MW as we explore next.

2.5.8. Implications for incompleteness around the Milky Way. While the MW and M31 profiles agree quite well out to 150 kpc, the MW appears to have a larger proportion of its satellite galaxies at small distances than both our simulations and M31. This could be a peculiarity of the MW profile, or it may be hinting at more satellites remaining to be discovered beyond 150 kpc from the MW. For example, the difference in shape could be due to the current presence of the LMC and the SMC near their pericenters around the MW (Kallivayalil et al., 2013). While Yniguez et al. (2014) found that potential incompleteness in the census of MW satellites meant there could be ∼10 classical dwarf satellite galaxies remaining to be discovered, which could bring the MW into better agreement with M31.

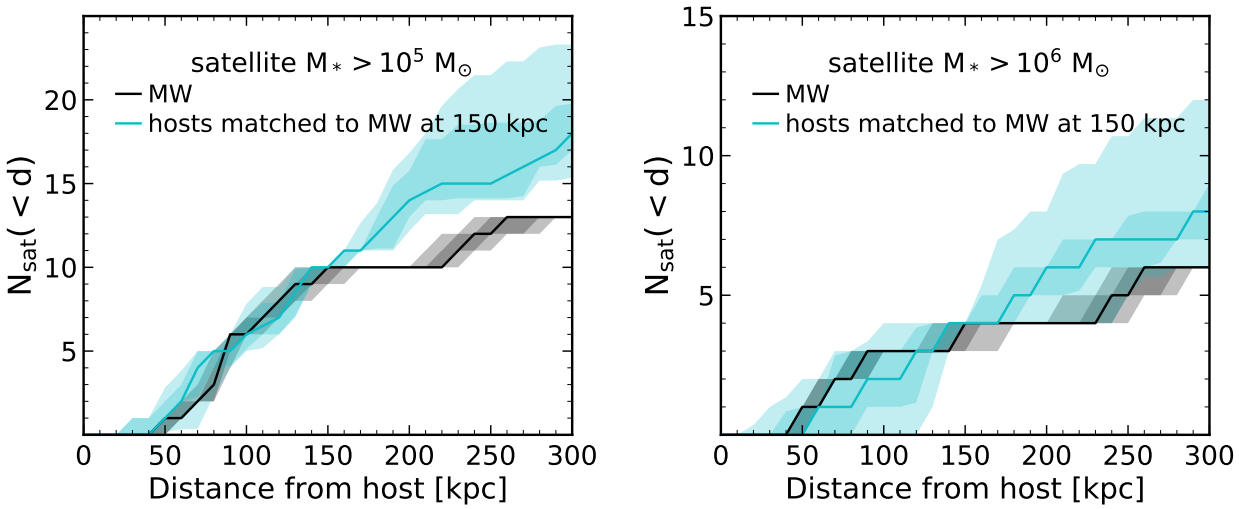

FIGURE 2.9. Comparison of the MW with simulated profiles matched to it at 150 kpc, and implications for incompleteness of satellite galaxies around the MW. Note that we compare our simulations to observations of the MW that have *not* been completeness-corrected. *Left:* The cumulative number of satellite galaxies with $M_* > 10^5$ $M_\odot$ as a function of 3D distance from the host for the MW (black) and simulations (blue) that match the number of satellites around the MW within 150 kpc. We find 8 profiles across all hosts and snapshots that meet this criteria. Within 150 kpc the agreement between the simulations and observations is excellent, but beyond 150 kpc all the simulations lie systematically at least 2 (and more commonly 5) satellites above the observations. This indicates that observations of the MW may be incomplete for satellites with $M_* > 10^5$ $M_\odot$. *Right:* Same as left, but for satellite galaxies with $M_* > 10^6$ $M_\odot$. We find 27 profiles across all hosts and snapshots that match the MW at this mass limit. The agreement between the simulations and observations spans the full distance range for this mass bin. Though the MW lies within simulation scatter, the simulation median is 1-2 satellites higher than the MW beyond 150 kpc. Observations of the MW are likely complete or nearly complete for satellite galaxies with $M_* > 10^6$ $M_\odot$ based on our simulations.

We expect observations of MW satellites to be complete down to at least $M_* \sim 10^5$ $M_\odot$ within 150 kpc and out of the plane of the disk. Beyond this distance and through the disk the completeness may be uncertain, as evidenced by the discovery of Antlia 2, which had been obscured by the MW disk. Here, we focus on implications for incompleteness without considering the effects of seeing through the MW's disk. While our theoretical

results are suggestive, a more in-depth account of observational completeness for 'classical' dwarf galaxies also depends on the surface brightness distribution of the population and their on-the-sky positions with respect to the Galactic plane (or any other foreground structure). Our simulations can provide more detailed predictions for these effects on the completeness of the satellite population, especially through the use of Gaia-like mocks (Sanderson et al., 2018), which we plan to pursue in future work.

To investigate potential incompleteness in observations of the MW's satellites (that have *not* been completeness-corrected), we examine how many additional satellites we would expect to find around the MW based on our simulations that match the MW profile out to 150 kpc. We choose simulated profiles for comparison by requiring them to have the same number of satellites within 150 kpc as the median value for the MW, which is 10 for $M_* > 10^5$ $M_\odot$. One host meets this criteria at four snapshots (m12z), and four hosts meet this criteria at a single snapshot each (m12w, m12r, Romeo, and Juliet), providing a total of 8 matched profiles.

Figure 2.9 (left) shows the range of simulated profiles that match the MW at 150 kpc compared to the observed MW profile, for satellites with $M_* > 10^5$ $M_\odot$. The simulations agree remarkably well with the MW below 150 kpc, which further strengthens our claim that if we match the profile at this distance, then we are accurately resolving survivability of satellites closer to the host. Notably, beyond $\sim$150 kpc the simulation profiles are systematically higher than the MW profile. In total, the simulation median profile has 5 more satellites than the MW median profile within 300 kpc. The lower 68 per cent (95 per cent) limits on the simulation profile imply that there may be at least 4 (2) more satellites at 150-300 kpc from the MW. If our simulations are representative of the real MW, then based on the median simulation profile, we predict that there should be 5 more satellites with $M_* > 10^5$ $M_\odot$ within 150-300 kpc of the MW.

We expect observational completeness to be better at higher satellite stellar masses, so we repeat this exercise for satellites with $M_* > 10^6$ $M_\odot$ to check if the agreement between simulations and observations is indeed better. At this satellite stellar mass threshold, the MW has 4 satellites within 150 kpc. We find that 8 out of the 12 simulated hosts match the MW's profile at 150 kpc for at least 1 snapshot out of 11, together providing a total of 27 matching profiles. Notably, m12i matches for 8 snapshots and Romeo matches for 5 snapshots.

Figure 2.9 (right) shows the range of simulated profiles that match the MW at 150 kpc compared to the observed MW profile, for satellites with $M_* > 10^6$ $M_\odot$. We find that the agreement between the simulations and observations spans the full range of distances in this satellite mass range. The 95 per cent simulation scatter almost completely encompasses the MW observational scatter below 150 kpc, and beyond that the 95 per cent simulation scatter overlaps with the upper half of the observational scatter. The lower 68 per cent (95 per cent) limits on the simulation profile imply that there may be at least 1 (0) more satellite with $M_* > 10^6$ $M_\odot$ to be discovered within 150-300 kpc of the MW. The median simulation profile indicates that there are on average 2 satellites in this mass range remaining to be discovered around the MW. Compared to the larger number of undiscovered satellites that we predict for the lower mass range, the observations of satellites with $M_* > 10^6$ $M_\odot$ appear to be more complete. We find that this strengthens our conclusion that the census of MW satellite galaxies may not be complete down to $M_* > 10^5$ $M_\odot$.

2.5.9. Incompleteness around M31. M31's satellite population is complete down to our lowest stellar mass limit ($M_* > 10^5$ $M_\odot$) and within 150 kpc of the host given the uniform depth and coverage of PAndAS in this area (McConnachie et al. 2009 and see Section 2.4 for more discussion). However, outside of the PAndAS footprint, the completeness limit for M31's satellite galaxies is not clear. We use our simulations as testing grounds to examine effects of this incompleteness on recovering M31's true radial profile. For simplicity and to

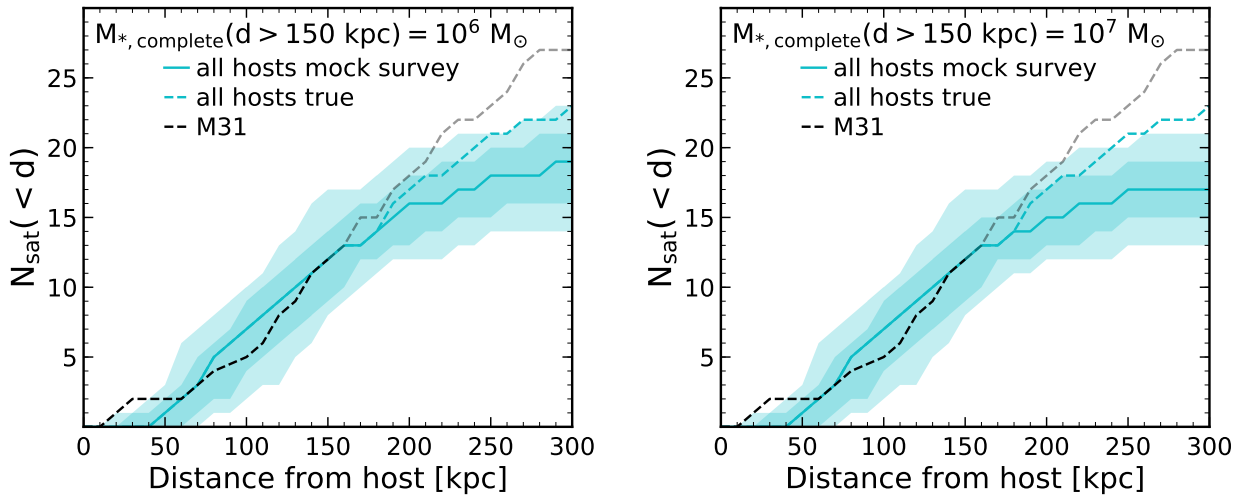

FIGURE 2.10. Implications for incompleteness of M31 satellites as a function of distance. M31's line is lighter where the observational data are known to be incomplete. *Left:* Solid blue line and scatter shows the radial profile measured by a mock survey that is complete to $M_* = 10^5$ $M_\odot$ within 150 kpc (to mock the PAndAS footprint) and complete to $M_* = 10^6$ $M_\odot$ for 150-300 kpc. Dashed blue line shows the true radial profile for all satellites with $M_* > 10^5$ $M_\odot$. Dashed black line shows M31's profile for comparison. Incompleteness causes the mock survey to miss ∼20 per cent of the satellites. *Right:* Same as left, but for a mock survey that is complete to $M_* = 10^7$ $M_\odot$ within 150-300 kpc. Here, incompleteness causes the mock survey to miss ∼25 per cent of the satellites. If our simulations are accurate representations of M31-like satellite populations, these results predict that there are 6-9 satellite galaxies to discover around M31.

match M31's profile (which has a median value of 27 satellites at 300 kpc), we select hosts from our simulations that have at least 20 (median over time) satellite galaxies within 300 kpc with $M_* > 10^5$ $M_\odot$: m12m, m12c, m12w, Juliet, and Louise. We perform a mock survey by selecting satellites in 2D projection along 1000 lines of sight. To mimic the PAndAS footprint, we assume that our mock observations are complete down to $M_* = 10^5$ $M_\odot$ within a projected radial distance of 150 kpc from the host, and within 150-300 kpc we assume two possible estimates of the completeness: $M_* > 10^6$ $M_\odot$ and $M_* > 10^7$ $M_\odot$.

Figure 2.10 shows the results of our mock surveys compared to the true radial profiles for the 5 hosts with M31-like profiles. Comparing our simulated true median profiles (blue

dashed) to the recovered profiles (blue solid, with shaded regions showing 68 per cent and 95 per cent scatter), we find that we typically recover 75-80 per cent (median) of our satellites, depending on the completeness mass. Thus, if our estimates of stellar completeness beyond 150 kpc are correct, M31 reasonably has 6-9 undetected satellites with $M_* > 10^5$ $M_\odot$ within 150-300 kpc of the host, which we obtain by applying 20-25 per cent incompleteness to M31's observed profile. It is also worth noting that beyond ∼200 kpc, the M31 profile lies above the scatter in the selected simulations. This may indicate that M31 is more massive than our simulated hosts, or that there is something else fundamentally different about M31 compared to our simulations. This result motivates deeper PAndAS-like surveys out to greater distances around M31, which are likely to find several dwarf galaxies, based on our simulations.

2.6. Summary and Discussion

Using the FIRE-2 baryonic cosmological zoom-in simulations of MW- and M31-mass halos, we study the radial profiles of satellite galaxies with $M_* > 10^5$ $M_\odot$. We explore 12 host-satellite systems: 8 isolated MW/M31-like galaxies from the Latte suite + m12z and 4 galaxies in LG-like pairs from the ELVIS on FIRE suite, where the hosts span $M_{200m} = 0.9-1.7\times10^{12}$ $M_\odot$. To reduce noise in profiles at small distances from satellites momentarily near pericenter, we time-average the simulated radial profiles over $z = 0-0.1$ (∼1.3 Gyr). We compare against the 3D profiles measured around the MW and M31 (including observational uncertainties in line-of-sight distance), and against the 2D profiles of MW analogs in the SAGA survey. Our main conclusions are as follows:

- The radial distributions of satellite galaxies with $M_* > 10^5$ $M_\odot$ within 300 kpc of their host in the FIRE-2 simulations agree well with LG observations. The scatter in the simulations spans the radial profiles of the MW and M31, and the median ratio of simulated-to-observed profiles is typically ∼1 for the MW and ∼1/2 for

M31. Though M31 has a relatively large satellite population, it is still within our simulation scatter.

- The radial concentration of the baryonic simulations generally agrees with LG observations, but the MW (and M31 in 2D projection) has a more concentrated shape than the simulations. If we examine simulations with the same number of satellites as the MW at d<150 kpc, we find excellent agreement with the MW down to ∼50 kpc. Beyond 150 kpc, the matched simulation profiles all lie above the MW profile. We predict 2-10 satellites (at 95%) with $M_* > 10^5$ $M_\odot$ to be discovered within 150-300 kpc from the MW.
- If we perform mock surveys with the same observational characteristics as PAndAS on our simulations, we recover on average 75-80 per cent of the true satellite population. Based on this, we predict there may be 6-9 undetected satellites around M31 and outside the PAndAS footprint depending on the (uncertain) completeness limit outside of the PAndAS footprint.
- 2D projected radial profiles of satellite galaxies with $M_* > 5 \times 10^6$ $M_\odot$ for the simulations also agree with the profiles for the 8 MW analogs from the SAGA survey. The scatter in the simulations spans a majority of SAGA profiles, though 3 SAGA hosts have fewer satellites at large distances (>100 kpc).
- The agreement we find in radial profiles does not depend strongly on *satellite* galaxy stellar mass. Thus, even at small distances (¡100 kpc) where satellite galaxies are subject to stronger tidal forces from the host's disk, our simulations resolve the survival and physical destruction of satellites down to our lower stellar mass limit ($M_* > 10^5$ $M_\odot$, with typical $M_{\rm peak} > 8 \times 10^8$ $M_\odot$ or $\sim 2 \times 10^4$ DM particles).
- Simulated *hosts* with larger stellar masses have fewer satellite galaxies at small distances ($\lesssim$ 100 kpc). We interpret this as caused by tidal destruction of satellite galaxies by the gravitational potential the host's disk. We find a similar correlation

with halo mass as well, which we interpret as a manifestation of more massive halos having bigger disks. We note, however, that we examined hosts only over a narrow host halo mass range $M_h = 0.9 - 1.7 \times 10^{12}$ $M_\odot$.

- The variation from host-to-host scatter among the different simulations dominates over time variation at large distances ($\gtrsim$100 kpc), while time variation is the dominant contributor to scatter at small distances ($\lesssim$50 kpc).
- KS tests between the radial profiles of the simulations the profiles of the MW and M31 show that most of the simulated profiles are consistent with being drawn from the same underlying distribution as LG observations. However, 4 (1) of the simulations have radial profiles inconsistent with the MW (M31).
- Consistent with previous studies, our dark matter-only simulations have many more subhalos at small distances (¡100 kpc), and hence larger concentrations in their radial profiles, than their baryonic counterparts. This corroborates the idea that the baryonic simulations have enhanced tidal destruction of satellites due to the additional disk potential present in baryonic hosts. We provide fits to the ratio of baryonic to dark matter-only subhalo counts as a function of distance, which one can use to renormalize existing DMO simulations to include baryonic effects.

We present a thorough comparison of satellite galaxy radial profiles around MW/M31-like galaxies in the FIRE-2 simulations to the LG and to MW analogs from the SAGA survey. Incorporating time dependence of the radial profile over the last 1.3 Gyr in the simulations is key to a robust comparison of the simulations with observations, because the profile at small distances from the host can be highly time-variable. Overall, we find that our simulations are generally representative of current observations. Specifically, combined with the recent results of Garrison-Kimmel et al. (2019a) and Garrison-Kimmel et al. (2019b), who analyzed the same FIRE simulation suite, we see broad agreement with the population of 'classical' dwarf galaxies ($M_* \gtrsim 10^5$ $M_\odot$) in the LG across a wide range of properties: stellar masses,

stellar velocity dispersion and dynamical mass profiles, star-formation histories, and now spatial distributions in terms of radial profiles. However, we emphasize that these simulations are not yet able to resolve ultra-faint dwarf galaxies, so it remains unclear how well simulations agree with the profiles of ultra-faints (especially in incorporating incompleteness). The simulations used here also do not yet include the most realistic treatment of cosmic ray physics implemented in the FIRE project, which has effects on the mass of the host and hence the survivability of satellites (Chan et al., 2018, Hopkins et al., in preparation).

Given the correlation with number of satellites at small distances with host stellar mass, we interpret the analogous dearth of subhalos in the baryonic simulations relative to the DMO simulations as primarily from tidal disruption of satellites by the baryonic disk. This agrees with a wealth of previous work that generally finds an excess of DMO subalos near the host relative to the number in baryonic simulations or DMO+analytical disk potential (Taylor & Babul, 2001; Hayashi et al., 2003; Read et al., 2006a,b; Berezinsky et al., 2006; D'Onghia et al., 2010; Peñarrubia et al., 2010; Brooks et al., 2013; Zhu et al., 2016; Errani et al., 2017; Garrison-Kimmel et al., 2017b; Sawala et al., 2017; Kelley et al., 2019).

The shape of the radial profile of satellite galaxies also has significant implications for how other satellite phenomena are measured. For example, the missing satellites problem (e.g. Moore et al., 1999; Klypin et al., 1999b) and the satellite plane problem (e.g. Pawlowski, 2018) are both sensitive to concentration of the radial profile, and the MW's satellite distribution is often found to be unusually concentrated compared to simulations (e.g. Zentner et al., 2005; Li & Helmi, 2008; Metz et al., 2009; Yniguez et al., 2014). However, controlling for the shape of the profile proves difficult because typical metrics of radial concentration do not necessarily produce the comprehensive description of spatial distribution that is needed to interpret observations. We find that DMO simulations have systematically higher radial concentration than baryonic simulations. Other studies have reached the same conclusion by

comparing DMO simulations to baryonic simulations or to DMO simulations with a semi-analytic model of galaxy formation (e.g. Kang et al., 2005; Ahmed et al., 2017). This suggests that DMO simulations alone cannot be used to accurately predict the shapes of observed radial profiles which are likely affected by baryonic processes.

We also find that while the radial concentration of the M31 profile agrees with our baryonic simulations, the MW is more concentrated than the baryonic simulations when we compare profile shape with R_{90}/R_{50}. The MW is more concentrated than the baryonic simulations (and even most of the DMO simulations) under this metric because 50 per cent of MW satellites are within 110 kpc of the MW, but the simulations only attain this fraction of satellites within $\gtrsim$140 kpc of the host on average. This is similar to what Yniguez et al. (2014) found by comparing the number of satellites within 100 and 400 kpc of their host for LG profiles and DMO simulations: the MW has a more concentrated shape than all of their simulations and M31. If we instead match the number of simulated satellites within 150 kpc of their host to the observed number within 150 kpc of the MW, we find that simulations meeting this criteria unanimously show a larger number of satellites within 150-300 kpc than the MW. We interpret this as potential evidence for incompleteness in the MW's satellite population at large distances, and our simulations predict there are on average 5 (at least 2) satellites with $M_* > 10^5$ $M_\odot$ to be discovered beyond 150 kpc from the MW.

Due to the peculiarity of the MW profile, we also use KS testing to accurately compare our simulations with observations. We find that all 12 of the simulated hosts have at least 10 snapshots matching M31's profile, and 9 of the hosts have at least 10 snapshots matching the MW's profile. We will examine the full three-dimensional spatial and dynamical distributions of satellite galaxies in detail and examine the satellite plane problem in our simulations in future work (Samuel et al., in preparation).

The spatial distribution of satellite galaxies correlates with attributes of the host galaxy, both in our simulations and in the LG. Importantly, the correlated host attributes are not

limited to the dark matter halo properties of the host, and the spatial distribution of satellites may be most strongly correlated with the host's baryonic features. DMO simulations are insensitive to the effects of a realistic host galaxy disk, and thus are not sufficient predictors of observed radial profiles at small distances which are the most influenced by the host's baryonic structure. We have provided a correction to such DMO radial profiles by modeling the depletion of subhalos by the baryonic disk as a function of distance from the host.

Acknowledgements

We thank Marla Geha, Risa Wechsler, and Ethan Nadler for their helpful comments. This research made use of Astropy,[4] a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013, 2018), the IPython package (Pérez & Granger, 2007), NumPy (Harris et al., 2020), SciPy (Jones et al., 2001), Numba (Lam et al., 2015), and matplotlib, a Python library for publication quality graphics (Hunter, 2007).

JS, AW, and SB were supported by NASA, through ATP grant 80NSSC18K1097 and HST grants GO-14734 and AR-15057 from the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. We performed this work in part at the Aspen Center for Physics, supported by NSF grant PHY-1607611, and at the KITP, supported NSF grant PHY-1748958. Support for SGK and PFH was provided by an Alfred P. Sloan Research Fellowship, NSF grant #1715847 and CAREER grant #1455342, and NASA grants NNX15AT06G, JPL 1589742, 17-ATP17-0214. Support for SRL was provided by NASA through Hubble Fellowship grant #HST-JF2-51395.001-A awarded by STScI. KE was supported by an NSF graduate research fellowship. MBK acknowledges support from NSF grant AST-1517226 and CAREER grant AST-1752913 and from NASA grants NNX17AG29G and HST-AR-13888, HST-AR-13896, HST-AR-14282, HST-AR-14554, HST-AR-15006, HST-GO-12914, and HST-GO-14191 from STScI. CAFG was supported by NSF through grants AST-1517491,

[4]http://www.astropy.org

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. JSB was supported by NSF AST-1518291, HST-AR-14282, and HST-AR-13888. We ran simulations using the Extreme Science and Engineering Discovery Environment (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.

CHAPTER 3

Planes of satellites around Milky Way/M31-mass galaxies in the FIRE simulations and comparisons with the Local Group

Published as Jenna Samuel, Andrew Wetzel, Sierra Chapman, Erik Tollerud, Philip F Hopkins, Michael Boylan-Kolchin, Jeremy Bailin, and Claude-André Faucher-Giguère in Monthly Notices of the Royal Astronomical Society, Volume 504, Issue 1, June 2021, Pages 1379–1397, `https://doi.org/10.1093/mnras/stab955`

3.1. Abstract

We examine the prevalence, longevity, and causes of planes of satellite dwarf galaxies, as observed in the Local Group. We use 14 Milky Way/Andromeda-(MW/M31) mass host galaxies from the FIRE-2 simulations. We select the 14 most massive satellites by stellar mass within $d_{\rm host} \leq 300\,{\rm kpc}$ of each host and correct for incompleteness from the foreground galactic disc when comparing to the MW. We find that MW-like planes as spatially thin and/or kinematically coherent as observed are uncommon, but they do exist in our simulations. Spatially thin planes occur in 1–2 per cent of snapshots during $z = 0 - 0.2$, and kinematically coherent planes occur in 5 per cent of snapshots. These planes are generally transient, surviving for < 500 Myr. However, if we select hosts with an LMC-like satellite near first pericenter, the fraction of snapshots with MW-like planes increases dramatically to $7 - 16$ per cent, with lifetimes of $0.7 - 3$ Gyr, likely because of group accretion of satellites. We find that M31's satellite distribution is much more common: M31's satellites lie within $\sim 1\sigma$ of the simulation median for every plane metric we consider. We find no significant difference in average satellite planarity for isolated hosts versus hosts in LG-like pairs. Baryonic

and dark matter-only simulations exhibit similar levels of planarity, even though baryonic subhaloes are less centrally concentrated within their host halos. We conclude that planes of satellites are not a strong challenge to ΛCDM cosmology.

3.2. Introduction

Astrometric measurements have revealed that a subset of the Milky Way (MW) satellite galaxies coherently orbit their host galaxy within a spatially thin plane ('thin' describes systems with minor-to-major axis ratios of $c/a \lesssim 0.3$, and 'coherent' indicates that a majority of satellites share the same orbital direction) (e.g., Lynden-Bell, 1976; Kroupa et al., 2005; Pawlowski et al., 2012a). Recently, precise proper motions from *Gaia* Data Release 2 have affirmed an even tighter orbital alignment of MW satellites than previously measured (Fritz et al., 2018; Pawlowski & Kroupa, 2020). Similar structures have also been observed around Andromeda (M31) (Ibata et al., 2013; Conn et al., 2013) and Centaurus A (Müller et al., 2018). However, the spatial and kinematic coherence of satellite planes beyond the Local Group (LG) is less certain because of projection effects, distance uncertainties, and the inaccessibility of proper motions. Even at the relatively close distance of M31, currently only two of its satellites have measured proper motions (Sohn et al., 2020), making it difficult to determine true 3D orbital alignment of the entire satellite population.

The cosmological significance of these satellite planes remains a topic of ongoing investigation, largely because of a lack of consensus on the incidence of planarity in both simulations and observations. Studies using dark matter-only (DMO) simulations have often yielded conflicting interpretations of how rare satellite planes are in the standard cosmological model of cold dark matter with a cosmological constant (ΛCDM). Most analyses of DMO simulations find such configurations to be rare, highly significant, and therefore possibly in conflict with ΛCDM (e.g. Metz et al., 2008; Pawlowski & McGaugh, 2014; Buck et al., 2016). However, DMO simulations combined with semi-analytic models of galaxy formation suggest that planes might be more common (Libeskind et al., 2009; Cautun et al., 2015), but this

is not a universal result (Pawlowski et al., 2014; Ibata et al., 2014b). Results from baryonic simulations have varied too, often relying on a much smaller sample of host-satellite systems compared to what is available from DMO simulations. Some baryonic simulations show evidence for a more natural presence of satellite planes in the universe (e.g. Libeskind et al., 2007; Sawala et al., 2016). While other baryonic results show that satellite planes can be uncommon, but find conflicting evidence for whether planes can be explained by anisotropic satellite accretion along filamentary structures (Ahmed et al., 2017; Shao et al., 2018, 2019).

Beyond just checking for the presence and significance of satellite planes in simulations, several authors have also explored what may cause planes to form, with mixed results. Though one might expect the host halo to affect satellite planes, Pawlowski & McGaugh (2014) found no connection between planes and host halo properties. Some authors have argued either for (Zentner et al., 2005; Libeskind et al., 2011) or against (Pawlowski et al., 2012b) the preferential infall of satellites along cosmic filaments as a causal factor in the formation of satellite planes. Li & Helmi (2008) proposed the accretion of satellites in small groups as an explanation of correlated orbits, and Wetzel et al. (2015a) showed that $25-50$ per cent of satellite dwarf galaxies in MW-mass hosts today previously were part of a group. Metz et al. (2007) even speculated that satellite planes arise naturally from the creation of tidal dwarf galaxies in fly-bys or mergers of larger galaxies.

Several authors have investigated the orbital stability of LG satellite planes. Recently, Riley & Strigari (2020) showed that globular clusters and stellar streams around the MW do not seem to be members of the satellite plane, suggesting that plane members may be recently accreted or in a particularly stable orbital configuration. Pawlowski et al. (2017) noted that integrating present-day satellite orbits either forward or backward in time typically leads to the disintegration of the plane, especially when sampling measurement uncertainties on satellite galaxy positions and velocities. Shaya & Tully (2013) took a different approach

and, by searching the dynamical parameter space of Local Volume satellites, found past trajectories that could possibly lead to the observed satellite planes.

Many previous attempts to investigate satellite planes have relied on simulations that may not resolve the dynamical evolution of "classical" ($M_* \geq 10^5\,\mathrm{M}_\odot$) dwarf galaxies, or that do not include baryonic physics. Insufficient resolution can lead to artificial satellite destruction (e.g. Carlberg, 1994; van Kampen, 1995; Moore et al., 1996; Klypin et al., 1999a; van Kampen, 2000; Diemand et al., 2007; Wetzel & White, 2010; van den Bosch & Ogiya, 2018). This may introduce a bias in satellite plane metrics if the destruction is spatially varying (such as near the host disc), and because earlier infalling satellites are preferentially destroyed, leading to an age bias that correlates with satellite orbit today (Wetzel et al., 2015a).

If baryonic effects act to create or destroy planes of satellites, then dark matter-only simulations may not be able to wholly capture the theoretical picture of satellite plane formation. The central disc in baryonic simulations tidally destroys satellites, altering their radial profile at small distances from the host (e.g., D'Onghia et al., 2010; Sawala et al., 2017; Garrison-Kimmel et al., 2017b; Nadler et al., 2018; Kelley et al., 2019; Rodriguez Wimberly et al., 2019; Samuel et al., 2020). This leads the surviving satellites to have more tangentially biased orbits (Garrison-Kimmel et al., 2017b, 2019a), but these effects do not necessarily imply an effect on planarity. In addition, Ahmed et al. (2017) found that the members of satellite planes in baryonic versus DMO simulations of the same host halo can be different, suggesting that baryonic effects may alter halo occupation in unexpected ways and hence affect satellite planes. Garrison-Kimmel et al. (2019a) also noted that satellites in baryonic simulations of LG-like pairs do not necessarily trace the most massive subhaloes in DMO runs of the same systems.

Outside of the MW, the satellite plane around M31 is somewhat more ambiguous. Taken as a whole, M31's satellites do not appear to be particularly planar, but a subset of 15

satellites lie within a significantly spatially thin plane and most of those are kinematically aligned, based on line-of-sight velocities (Conn et al., 2013; Ibata et al., 2013). Many works have focused in on this particular subset, but it is important to understand the overall satellite distribution, because there are no clear evolutionary differences between M31 plane members and non-members (Collins et al., 2015).

Satellite planes outside of the LG are more difficult to robustly characterize because of projection effects and larger distance uncertainties. Studies using the Sloan Digital Sky Survey (SDSS) database have revealed that while there is evidence for spatial flattening of satellites (e.g., Brainerd, 2005), their kinematic distribution is unlikely to indicate a coherently orbiting satellite plane (Phillips et al., 2015). Furthermore, the Satellites Around Galactic analogues (SAGA) survey (Geha et al., 2017), which aims to study satellites of ~ 100 MW analogues in the nearby Universe, has found little evidence for coherently orbiting satellite planes (Mao et al., 2021).

In this paper, we seek to understand if the FIRE-2 simulations contain satellite planes similar to those found in the Local Group, whether those satellite planes are long-lived or transient, and if the presence of satellite planes correlates with host or satellite properties. We leave comparisons to systems outside of the LG for future work. We organize this paper as follows: in Section 3.3 we describe our simulations and satellite selection criteria, in Section 3.4 we describe the 3D positions and velocities of Local Group satellites used, in Section 3.5 we describe the plane metrics we apply to simulations and observations, in Section 3.6 we present our results of planarity in simulations compared to observations, and in Section 3.7 we discuss our conclusions and their implications for observed satellite planes.

3.3. Simulations

The zoom-in simulations we use in this work reproduce the mass functions, radial distributions, and star formation histories of classical ($M_* \geq 10^5\,\mathrm{M}_\odot$) dwarf galaxies around

MW/M31-like hosts (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a,b; Samuel et al., 2020).

We use two suites of cosmological zoom-in hydrodynamic simulations from the Feedback In Realistic Environments (FIRE) project[1]. Latte is currently a suite of 7 isolated MW/M31-mass galaxies with halo masses $M_{200m} = 1 - 2 \times 10^{12}\,M_\odot$[2] introduced in Wetzel et al. (2016). We selected the Latte halos for zoom-in re-simulation from a periodic volume dark matter simulation box of side length 85.5 Mpc. We selected two of the Latte halos (m12r and m12w) to host an LMC-mass subhalo at $z = 0$ within their initial DMO simulations, though after re-simulation with baryonic physics the orbital phase of these subhaloes changes and they are no longer near pericenter (Samuel et al., 2020). Latte gas and star particles have initial masses of $7070\,M_\odot$, but at $z = 0$ a typical star particle has mass $\approx 5000\,M_\odot$ because of stellar mass loss. Dark matter particles have a mass resolution of $m_{dm} = 3.5 \times 10^4\,M_\odot$. The gravitational softenings (comoving at $z > 9$ and physical at $z < 9$) of dark matter and stars particles are fixed: $\epsilon_{dm} = 40$ pc and $\epsilon_{star} = 4$ pc (Plummer equivalent). The gas softening is fully adaptive, matched to the hydrodynamic resolution, and the minimum gas resolution (inter-element spacing) and softening length reached in Latte is ≈ 1 pc. We also use an additional simulation of an isolated MW/M31-mass galaxy (m12z), simulated at higher mass resolution ($m_{baryon,ini} = 4200\,M_\odot$).

The second suite of simulations we use is "ELVIS on FIRE". This suite consists of three simulations, containing two MW/M31-mass galaxies each, wherein the main halos were selected to mimic the relative separation and velocity of the MW-M31 pair in the LG (Garrison-Kimmel et al., 2014, 2019a,b). ELVIS on FIRE has $\approx 2\times$ better mass resolution than Latte: the Romeo & Juliet and Romulus & Remus simulations have $m_{baryon,ini} = 3500\,M_\odot$ and the Thelma & Louise simulation has $m_{baryon,ini} = 4000\,M_\odot$.

[1] `https://fire.northwestern.edu/`

[2] '200m' indicates a measurement relative to 200 times the mean matter density of the Universe

We ran all simulations with the upgraded FIRE-2 implementations of fluid dynamics, star formation, and stellar feedback (Hopkins et al., 2018). FIRE uses a Lagrangian meshless finite-mass (MFM) hydrodynamics code, GIZMO (Hopkins, 2015). GIZMO enables adaptive hydrodynamic gas particle smoothing depending on the density of particles while still conserving mass, energy, and momentum to machine accuracy. Gravitational forces are solved using an upgraded version of the N-body GADGET-3 Tree-PM solver (Springel, 2005).

The FIRE-2 methodology includes detailed subgrid models for gas physics, star formation, and stellar feedback. Gas models used include: a metallicity-dependent treatment of radiative heating and cooling over $10 - 10^{10}$ K (Hopkins et al., 2018), a cosmic ultraviolet background with early HI reionization ($z_{\rm reion} \sim 10$) (Faucher-Giguère et al., 2009), and turbulent metal diffusion (Hopkins, 2016; Su et al., 2017; Escala et al., 2018). We allow gas that is self-gravitating, Jeans-unstable, cold ($T < 10^4$ K), dense ($n > 1000$ cm^{-3}), and molecular (following Krumholz & Gnedin (2011)) to form stars. Star particles represent individual stellar populations under the assumption of a Kroupa stellar initial mass function (Kroupa, 2001). Once formed, star particles evolve according to stellar population models from STARBURST99 v7.0 (Leitherer et al., 1999). We model several stellar feedback processes including core-collapse and Type Ia supernovae, continuous stellar mass loss, photoionization, photoelectric heating, and radiation pressure.

For all simulations, we generate cosmological zoom-in initial conditions at $z = 99$ using the MUSIC code (Hahn & Abel, 2011), and we save 600 snapshots from $z = 99$ to 0, with typical spacing of $\lesssim$25 Myr. All simulations assume flat ΛCDM cosmologies, with slightly different parameters across the full suite: $h = 0.68 - 0.71$, $\Omega_\Lambda = 0.69 - 0.734$, $\Omega_m = 0.266 - 0.31$, $\Omega_b = 0.0455 - 0.048$, $\sigma_8 = 0.801 - 0.82$, and $n_{\rm s} = 0.961 - 0.97$, broadly consistent with Planck Collaboration et al. (2018).

3.3.1. Halo finder. We use the ROCKSTAR 6D halo finder (Behroozi et al., 2013a) to identify dark matter halos and subhaloes in our simulations. We include a halo in the catalog if its bound mass fraction is > 0.4 and if it contains at least 30 dark matter particles within a radius that encloses 200 times the mean matter density, R_{200m}. We generate a halo catalog for each of the 600 snapshots of each simulation, using only dark matter particles. The subhaloes that we use in this work (within 300 kpc of their host) are uncontaminated by low-resolution dark matter particles. We then construct merger trees using CONSISTENT-TREES (Behroozi et al., 2013b).

We describe our post-processing method for assigning star particles to (sub)halos further in Samuel et al. (2020). First, we identify all star particles within 0.8 R_{halo} (out to a maximum 30 kpc) of a halo as members of that halo. Then, we further clean the member star particle sample by selecting those (1) that are within 1.5 times the radius enclosing 90 per cent of the mass of member star particles (R_{90}) from both the center-of-mass position of member stars and the dark matter halo center, and (2) with velocities less than twice the velocity dispersion of member star particles (σ_{vel}) with respect to the center-of-mass velocity of member stars. We iterate through steps (1) and (2) until the total mass of member star particles (M_*) converges to within 1 per cent. Finally, we save halos for analysis that contain at least 6 star particles and that have an average stellar density $> 300\,M_\odot\,kpc^{-3}$. We performed this post-processing and the remainder of our analysis using the `GizmoAnalysis` and `HaloAnalysis` software packages (Wetzel & Garrison-Kimmel, 2020a,b).

3.3.2. Satellite selection. Throughout this paper we refer to the central MW/M31-mass galaxies in our simulations as hosts, and their surrounding population of dwarf galaxies within 300 kpc as satellites. Our host galaxies have stellar masses in the range $M_* \sim 10^{10-11}\,M_\odot$ and dark matter halos in the mass range $M_{200m} = 0.9 - 1.7 \times 10^{12}\,M_\odot$. The eight Latte+m12z simulations contain a single isolated host per simulation. Each of the three ELVIS on FIRE simulations contains two hosts in a LG-like pair, surrounded by their

own distinct satellite populations. Thus, we use a total of 14 host-satellite systems to study satellite planes in this work. Our fiducial redshift range is $z = 0-0.2$ (114 snapshots), giving us a time baseline of ~ 2.4 Gyr over which to examine the presence of satellite planes at late times in our simulations. We present our results treating each snapshot as a separate (but not fully independent) realization and stacking snapshots across hosts. This allows us to mitigate the time-variability and host-to-host scatter in the satellite distribution at small distances from the host, and achieve robust comparisons of simulations and observations. We also consider a longer time window ($z = 0-0.5$, 219 snapshots, ~ 5.1 Gyr) in Section 3.6.2.2 in order to examine the lifetimes of planar structures and the coincidence of spatial thinness and kinematic coherence in our simulations.

We consider two ways to select simulated satellite galaxies for comparison to the MW. Our primary method is to select a fixed number of satellites around each host, by choosing the 14 satellites with highest stellar mass from our simulations, to match the number of observed MW satellites that have $M_* \geq 10^5\,\mathrm{M}_\odot$. We also choose the 15 most massive satellites around hosts for our comparison to M31 (see Section 3.6.1.2 for more details). Satellites with $M_* \geq 10^5\,\mathrm{M}_\odot$ contain ≥ 20 star particles and have peak halo masses of $M_{\rm peak} \geq 8\times10^8\,\mathrm{M}_\odot$ ($\gtrsim 2.3\times10^4$ dark matter particles prior to infall). Satellite galaxies with $M_* \geq 10^5\,\mathrm{M}_\odot$ are also nearly complete in observations (e.g. Koposov et al., 2007; Tollerud et al., 2008; Walsh et al., 2009; Tollerud et al., 2014; Martin et al., 2016), so we choose this as our nominal stellar mass limit to select satellites around the MW and M31. As an example, at $z = 0$, the satellite with the lowest stellar mass in our fixed-number satellite selection criteria has $M_* = 5.6\times10^4\,\mathrm{M}_\odot$ (11 star particles), which is enough to at least indicate the presence of a true satellite, given that it also satisfies the subhalo criteria outlined in Section 3.3.1.

We also consider a stellar mass threshold selection method in Section 3.6.3.2 whereby we require satellites to have $M_* \geq 10^5\,\mathrm{M}_\odot$ and maintain the same distance cutoff ($d_{\rm host} \leq$

300 kpc). This selection means that the number of satellites considered around all hosts varies from 10 to 31 in the redshift range $z = 0 - 0.2$. See Samuel et al. (2020) for more details on the radial distributions and resolution of simulated satellites meeting our criteria, and completeness estimates in the Local Group. See Garrison-Kimmel et al. (2019a,b) for how the stellar mass, velocity dispersion, dynamical mass, and star-formation histories of satellite dwarf galaxies in our simulations all broadly agree with MW and M31 observations, making these simulations compelling to use to examine planarity.

3.4. Observations

We consider all known MW satellite galaxies with $M_* \geq 10^5\,\mathrm{M}_\odot$ and $d_{\mathrm{host}} \leq 300\,\mathrm{kpc}$, based on the satellite stellar masses and galactocentric distances listed in Table A1 of Garrison-Kimmel et al. (2019a). While we are not confident that our halo finder is able to correctly identify analogues of the Sagittarius dwarf spheroidal (Sgr I) galaxy, given its significant tidal interactions, we include it in our observational sample, because it is a historical member of the MW's satellite plane. Excluding Sgr I from the MW satellite galaxy sample does not significantly change the resulting spread in the MW's plane metrics, and therefore we achieve essentially the same results in our comparisons to simulations regardless of this choice. For each observed satellite, we take the sky coordinates and heliocentric distances with uncertainties from McConnachie (2012). Furthermore, we include Crater II and Antlia II, which meet our stellar mass and distance criteria as described in Samuel et al. (2020), and use the positions and uncertainties from their discovery papers (Torrealba et al., 2016, 2018). This brings the total number of MW satellites that we consider in this study to 14. We consider effects of observational incompleteness from the Galactic disc in Section 3.6.3.1.

We use proper motions from *Gaia* Data Release 2 as presented in Fritz et al. (2018). We use the larger of the statistical or systematic uncertainties on *Gaia* proper motions, which typically is the systematic uncertainties. We take line-of-sight heliocentric velocities (v_{los})

for MW satellites and their uncertainties from Pawlowski & Kroupa (2020) and Fritz et al. (2018), where available. To supplement this, we use the proper motions and $v_{\rm los}$ for the Magellanic Clouds presented in Kallivayalil et al. (2013), and Antlia II's kinematics come from its discovery paper (Torrealba et al., 2018).

In our analysis of the MW satellite plane, we first sample the heliocentric distances, line-of-sight velocities, and proper motions 1000 times assuming Gaussian distributions on the uncertainties. We then convert these values to a Cartesian galactocentric coordinate system using Astropy (Astropy Collaboration et al., 2013, 2018). We measure planarity on the resulting satellite phase space coordinates in the same way we describe for simulated satellites in Section 3.5.

We take a different approach to sample M31's satellites. We impose the same stellar mass limit of $M_* \geq 10^5\,{\rm M}_\odot$ and 3D distance limit of $d_{\rm host} \leq 300\,{\rm kpc}$, but we additionally require that the projected distance from M31 listed in McConnachie (2012) adhere to $d_{\rm host,proj} \leq 150\,{\rm kpc}$, because M31's satellite population is most complete within this range from the Pan-Andromeda Archaeological Survey (PAndAS, McConnachie et al., 2009) coverage. We sample 1000 line-of-sight distances for each satellite, using the posterior distributions published in Conn et al. (2012) where available, and elsewhere assuming Gaussian distributions on distance uncertainties (McConnachie, 2012; Martin et al., 2013a). We assume that M32 and NGC205 have the same posterior distance distribution as M31 itself because they are too close to M31 to reliably determine their line-of-sight distances. The double-peaked posteriors of AndIX and AndXXVII cause the actual number of satellites within $d_{\rm host} \leq 300\,{\rm kpc}$ of M31 in each sample to range from 14 to 16, but this is unlikely to cause significant differences in our analysis. We take the line-of-sight velocities for M31 satellites from McConnachie (2012); Tollerud et al. (2012); Collins et al. (2013), and we use them for the 2D kinematic coherence metric described in Section 3.5.

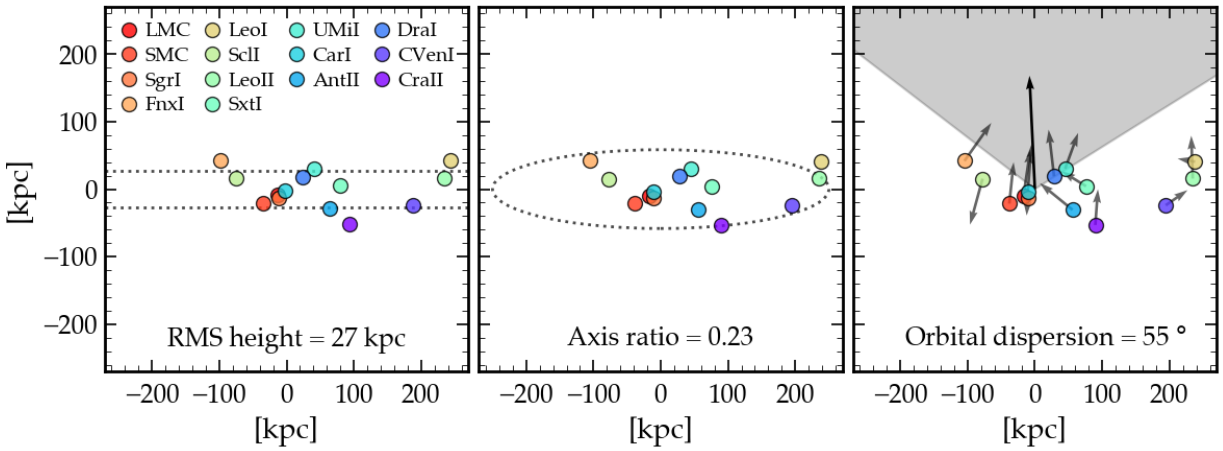

FIGURE 3.1. Diagram showing each plane metric that we use, as measured on the 3D positions and velocities of 14 MW satellites ($M_* \geq 10^5\,\mathrm{M}_\odot$ and $d_{\rm host} \leq 300\,\mathrm{kpc}$), shown in order of decreasing stellar mass. All planes are centered on the MW and observational uncertainties are neglected here for visual clarity. RMS height ($\Delta_{\rm h}$, left) is the root-mean-square distance of satellites from the satellite midplane. Axis ratio (middle) is the ratio of the minor-to-major axes (c/a) from the moment of inertia tensor of satellite positions. The ellipse shown has the same minor-to-major axis ratio as the MW's satellites. Orbital pole dispersion ($\Delta_{\rm orb}$, right) is the root-mean-square angle in the range $[0°, 360°]$ of the angular momentum unit vectors of satellites around their average direction. We show each metric in the same projection, to illustrate that the MW's satellite plane is kinematically coherent *within* a spatially thin plane.

3.5. Methods

Figure 3.1 is a visual demonstration of how we measure planarity using two spatial metrics and one kinematic metric. We show these metrics as measured on the MW's 14 satellites with $M_* \geq 10^5\,\mathrm{M}_\odot$ and $d_{\rm host} \leq 300\,\mathrm{kpc}$. For clarity we do not show the effects of observational uncertainties here, which have the largest effect on kinematic coherence, but we do include them in our analysis. Our planarity metric definitions are based on and consistent with those from e.g., Cautun et al. (2015); Pawlowski et al. (2015); Pawlowski & Kroupa (2020). We require all planes to pass through the center of the host galaxy. Below, we describe in detail each metric and how we calculated it at each simulation snapshot.

3.5.1. Spatial metrics of planarity. We measure the spatial coherence of satellite galaxies in two ways: root-mean-square (RMS) height ($\Delta_{\rm h}$) and minor-to-major axis ratio (c/a). The RMS height of a satellite distribution characterizes the vertical spread of satellites above and below a plane using the RMS component of satellites' 3D positions along the direction normal to a plane according to Equation 3.1. This can be thought of as the thickness or height of the plane. We randomly generate 10^4 planes centered on the host galaxy and quote the minimum value amongst these iterations.

$$\Delta_{\rm h} = \sqrt{\frac{\sum_{\rm i=1}^{\rm N_{sat}} (\hat{\mathbf{n}}_{\perp} \cdot \vec{x}_{\rm i})^2}{\rm N_{sat}}} \tag{3.1}$$

We also use the minor-to-major axis ratio (c/a) of the satellite spatial distribution to characterize spatial planes with a dimensionless metric. This is the ratio of the square root of the eigenvalues of the inertia tensor corresponding to the minor (c) and major (a) axes. We define a modified moment of inertia tensor treating satellites as unit point masses, weighting each one equally regardless of its stellar or halo mass, so it is a purely geometrical measure of the satellite distribution. The elements of the 3D inertia tensor are given by Equation 3.2.

$$I_{ij} = \sum_{k=1}^{N_{\rm sat}} \sum_{\alpha=1}^{3} \delta_{ij} r_{\alpha,k}^2 - r_{\alpha i,k} r_{\alpha j,k} \tag{3.2}$$

We explored a third metric of spatial planarity, enclosing angle, motivated by the desire to mitigate effects of radially concentrated satellite distributions on planarity measurements. We define enclosing angle as the smallest angle that encompasses the population of satellites, as measured off of the 'midplane' of the satellite plane. Similar to the galactocentric latitude (b_c) used in Section 3.6.3.2, the coordinate origin is placed at the center of the host galaxy. Enclosing angle ranges from 0 to 180 degrees by definition, where a measured angle of near 180 degrees indicates an isotropic distribution of satellites. Similar to the method used for

RMS height, in practice we randomly orient planes centered on the host galaxy from which to measure enclosing angle, and find the minimum angle from these iterations. We found that this metric was significantly noisier over time compared to the other spatial metrics, and often selected a different plane orientation from RMS height and axis ratio, so we do not use it in our final analysis.

3.5.2. Kinematic metrics of planarity. We consider both 3D and 2D measures of orbital kinematic coherence of satellite populations to compare against observed 3D velocities of satellites in the MW, and line-of-sight velocities ($v_{\rm los}$) of satellites around M31. The 3D metric we use is orbital pole dispersion ($\Delta_{\rm orb}$), which describes the alignment of satellite orbital angular momenta relative to the average satellite orbital angular momentum vector for the entire satellite population. We are not taking into account the magnitude of satellite orbital velocities, so orbital pole dispersion is a measure of purely directional coherence in satellite orbits. The orbital pole dispersion is defined as the RMS angular distance of the satellites' orbital angular momentum vectors with respect to the population's average orbital angular momentum direction, and is given by Equation 3.3. A system with all satellite orbital angular momenta aligned will have $\Delta_{\rm orb} = 0°$, while a random, isotropic distribution of satellite velocities has $\Delta_{\rm orb} \sim 180°$.

$$\Delta_{\rm orb} = \sqrt{\frac{\sum_{i=1}^{\rm N_{sat}} [\arccos(\hat{\mathbf{n}}_{\rm orb,avg} \cdot \hat{\mathbf{n}}_{\rm orb,i})]^2}{\rm N_{sat}}} \tag{3.3}$$

To investigate 2D orbital kinematic coherence around M31 we examine whether satellites share the same 'sense of orbital direction' around their host galaxy. We measure this by computing the maximum fraction ($f_{v_{\rm los}}^{\rm max}$) of satellites with opposing (approaching or receding) $v_{\rm los}$ on the left and right 'sides' of a satellite distribution. A fraction close to unity indicates a highly coherent system, and a fraction of 0.5 represents a purely isotropic system. We

compute this fraction along 10^3 randomly generated lines of sight in the simulations, and use the full distribution to compare to M31 as described in Section 3.6.1.2.

3.5.3. Statistically isotropic realizations of satellite positions and velocities. To compare the 'true' satellite planes (as measured at each snapshot) across different simulations, we quantify the likelihood of measuring thinner or more kinematically coherent planes in a statistically isotropic distribution of satellites. This is a more general characterization of planarity, independent of the actual values measured for observed systems, that can also address whether satellite planes are statistically significant. We generate isotropic realizations of satellite positions by randomly generating 10^4 polar and azimuthal angles for each satellite, keeping their radial distance from the host fixed, following Cautun et al. (2015). For isotropic kinematic distributions, we generate random unit velocities (using a similar prescription as for the randomization of angular coordinates) while also randomizing the angular spatial coordinates of each satellite. We then measure planarity for each of the 10^4 realizations. We quantify the significance of a planar alignment by quoting the fraction ($f_{\rm iso}$) of isotropic realizations with smaller values of plane metrics than the true value at each snapshot. In effect this is the conditional probability of finding a more planar distribution of satellites among the isotropic realizations. A fraction $f_{\rm iso} \leq 0.5$ indicates that the true satellite distribution is more planar than a statistically isotropic distribution of satellites, and we define $f_{\rm iso} \leq 0.05$ to mean the true satellite distribution is significantly planar.

3.6. Results

3.6.1. Comparisons of simulations and the Local Group. As we showed in Samuel et al. (2020), the simulations are a reasonable match to the radial distribution of satellites in the LG as a function of both distance from the host and stellar mass of the satellite. This provided an important first benchmark of just the 1D radial positions of satellites in our simulation. We now seek to leverage the full 3D positions and velocities of satellites

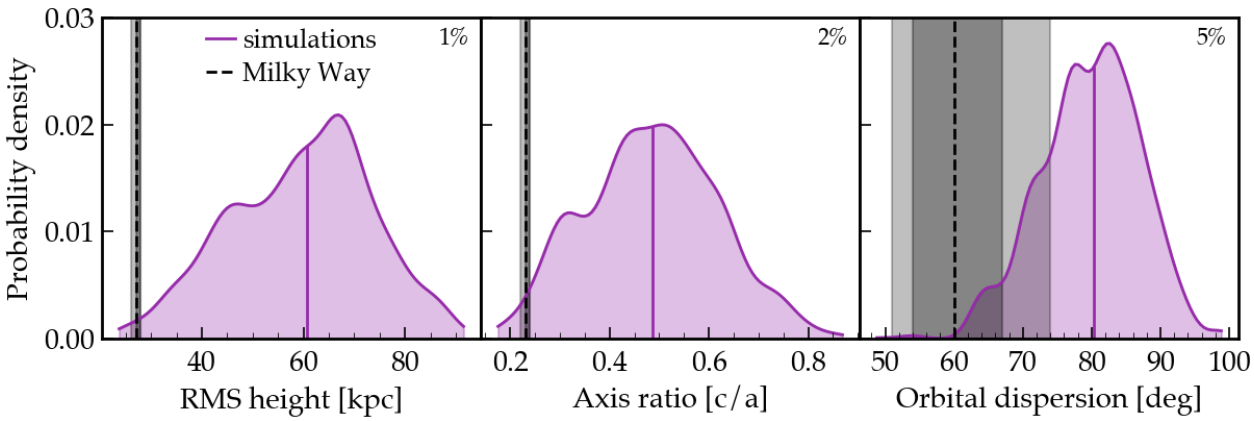

FIGURE 3.2. Planarity of simulated satellite galaxies ($N_{\rm sat} = 14$ and $d_{\rm host} \leq 300\,{\rm kpc}$) around MW/M31-mass hosts compared to the MW's satellite plane. We model incompleteness in the simulations by excluding any satellites that lie within $\pm 12^\circ$ of the plane of the host galaxy's stellar disc. We generate KDEs (purple) using 114 snapshots over $z = 0 - 0.2$ for each of the 14 simulated hosts, and the solid vertical colored lines are the distribution medians. We show MW observations (black) for 14 satellites with 68 (95) per cent spread from observational uncertainties. The number in the top right of each panel is the per cent of snapshots that are MW-like, which lie at or below the MW upper 68 per cent limit. For all metrics we consider, we find some ($1 - 5$ per cent) snapshots that are at least as planar as the MW, though they are rare.

in our simulations (and around the MW) to characterize satellite planes. We compare our simulations to observations of LG satellites, leaving comparisons to systems such as other MW/M31 analogues and Centaurus A for future work. In this section, we make physically rigorous comparisons using mock observations that include disc completeness corrections. In subsequent sections we further explore selection effects on measured satellite planes and possible physical origins of satellite planes.

3.6.1.1. *MW-like planes.* We select the 14 most massive satellites in M_* within $d_{\rm host} \leq 300\,{\rm kpc}$ to compare planarity in simulations and the 14 MW satellites in our observational sample. Furthermore, we apply a simple completeness correction for seeing through the MW's disc by first excluding all satellites that lie within a galactocentric latitude of $|b_c| \leq 12^\circ$ from the host's galactic disc (Pawlowski, 2018), and then choosing the 14 most massive

satellites from the remaining population. See Section 3.6.3.2 for an investigation of how disc incompleteness affects planarity metrics.

Figure 3.2 shows plane metrics for simulated satellites stacking over 114 snapshots spanning $z = 0 - 0.2$, compared to the MW satellite plane. Spatial plane metrics for the MW are tightly constrained by well-measured 3D positions of MW satellites. The MW's satellite plane is thinner and more kinematically coherent than most of our simulated satellite systems. We define MW-like planes as those with plane metrics at or below the one sigma upper limit on the MW's corresponding distribution. Notably, the MW's plane is significantly spatially flattened compared to the average simulation when measured by RMS height and axis ratio.

While MW-like spatial planes are rare in our simulations, we do identify satellite populations that are as thin as the MW's plane in $1 - 2$ per cent of our full sample. We compute each plane metric independently, but we discuss instances of satellite planes that are simultaneously both thin and kinematically coherent in Section 3.6.2.2. The occurrence of thin planes in $1 - 2$ per cent of snapshots holds over both our fiducial time baseline of $z = 0 - 0.2 \approx 2.4$ Gyr (114 snapshots per host, 1,596 snapshots in total) and also over the longer interval $z = 0 - 0.5 \approx 5.1$ Gyr (219 snapshots per host, 3,066 snapshots in total), an indication of the robustness of the measurement.

The uncertainties in 3D velocities of MW satellites are much larger than the uncertainties in their 3D positions, and this leads to a much wider spread in orbital pole dispersion of the MW compared to the spatial metrics. However, the MW's satellites still have highly correlated orbits relative to the simulations, with only 5 per cent of the simulations having a plane at least as kinematically coherent as the MW's upper one sigma limit during $z = 0-0.2$. The fraction of the full sample containing these planes actually increases to 8 per cent when measured over $z = 0 - 0.5$, likely from the correlated infall of satellites in groups or along filaments at earlier times. The spread in the MW's orbital pole dispersion is large compared

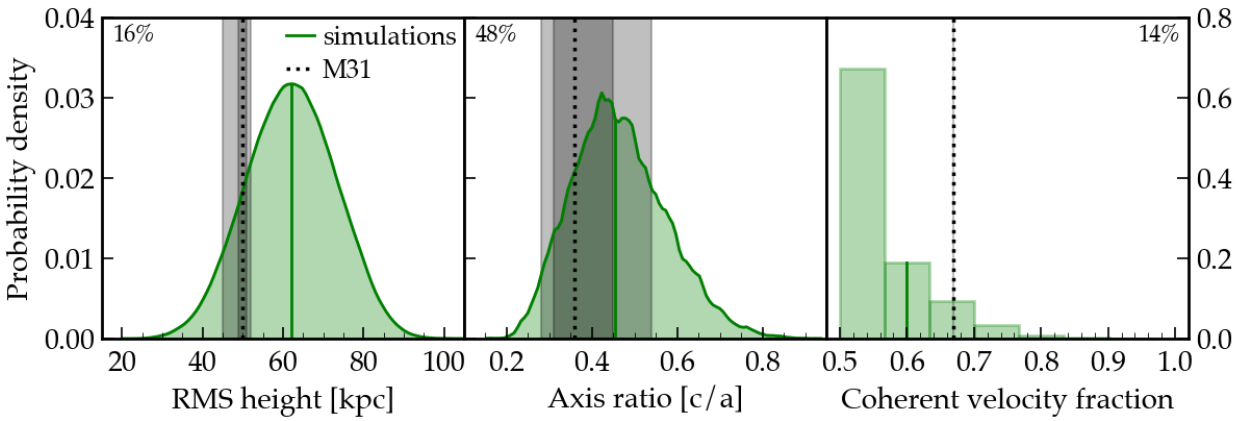

FIGURE 3.3. Planarity of the 15 simulated satellite galaxies with the highest stellar mass within $d_{\rm host,proj} \leq 150\,{\rm kpc}$ of each MW/M31-mass host (consistent with completeness in PAndAS). We generate KDEs (green) using 114 snapshots over $z = 0 - 0.2$ for each of the 14 simulated hosts. The solid vertical colored lines are the medians of each distribution. We show the M31 data (black) for 16 satellites with 68 (95) per cent spread in plane metrics from line-of-sight (LOS) distance uncertainties. The number in the top of each panel is the per cent of snapshots that are at least as planar as the M31 upper 68 per cent limits. Selected in this general way, the simulations are about as planar as M31. The average RMS height (left) of simulations is somewhat thicker than M31's satellite population as a whole, but M31 is still within $\sim 1\sigma$ of the simulation peak. Typical simulation axis ratios (center) are even more similar to M31's satellites. In the right panel, more planar snapshots are shown to the right of the M31 value. LOS velocity uncertainties are too small to broaden the M31 velocity coherence measurement M31's satellites are slightly more kinematically coherent than most simulations, but only by $\sim 1\sigma$, consistent with the spatial planarity comparisons.

to the spatial metrics, so we also provide the fraction of the simulation sample lying at or below the median MW value, 0.3 per cent. There are even a few (5) snapshots that extend below the MW distribution.

The MW's satellite kinematics, while rare, do not appear to be extreme outliers compared to our simulations. This broadly agrees with Pawlowski & Kroupa (2020), who found that $\sim 2 - 3$ per cent of hosts at $z = 0$ in the IllustrisTNG simulations (Pillepich et al., 2018; Nelson et al., 2019) have satellites as orbitally aligned as the MW. However, the comparison between our work and theirs is not one-to-one: they vary the number of satellites included in

plane calculations ($N_{\rm sat} = 3-11$) in both simulations and observations in order to account for the "look elsewhere" effect (the spurious detection of high significance events from searching a large parameter space), but they find that their conclusions do not vary for any number of plane members greater than three. The IllustrisTNG simulations they use allow them to analyze a larger number of hosts, in part because they choose to include dark subhaloes as satellites in order to maximize their sample size of hosts with at least 11 satellites. The larger host sample size comes at the cost of resolution though, with $m_{\rm DM} = 7.5 \times 10^6\,{\rm M}_\odot$, $m_{\rm baryon} = 1.4 \times 10^6\,{\rm M}_\odot$, and $\epsilon_{\rm DM,*} = 0.74$ kpc.

In contrast, our planarity metrics are predicated on matching the number of observed satellites ($N_{\rm sat} = 14$) and we only have 14 hosts. Instead, we leverage our time resolution to increase our sample size given that our planes are often transient features (see Section 3.6.2.2). Our simulations also have order-of-magnitude higher resolution, which may allow planes of satellites to survive that would be disrupted in lower resolution simulations. This is evidenced by their broad agreement with the MW and M31 in their radial distributions down to ~ 50 kpc (Samuel et al., 2020). Our measured plane metrics should be considered upper limits on absolute planarity at each snapshot. If we instead varied $N_{\rm sat} = 3-14$, to test for the look-elsewhere effect, we might find even thinner or more coherent planes. Likewise, our quoted fractions of MW-like planes are upper limits on the incidence of MW-like planarity, as this can only be diminished by accounting for the look-elsewhere effect. Because we are always choosing a larger number of plane members ($N_{\rm sat} = 14$) than used by Pawlowski & Kroupa (2020), which yields larger values of plane metrics in our case, we compare just the fractions of our samples that are MW-like instead of absolute plane metrics.

As a caveat to these kinematic results, we note that using a slightly different proper motion sample for the observed MW satellites leads to a reduced fraction of snapshots with MW-like kinematic planes. If we adopt the 'best-available' observed proper motions from Pawlowski & Kroupa (2020), the fraction of snapshots having a plane at least as kinematically

coherent as the MW's upper one sigma limit during $z = 0 - 0.2$ decreases to 0.3 per cent (see Figure B.1). However, this different proper motion data set does not qualitatively change any of our other results in the following sections, and it has no effect on the measured spatial planarity in simulations or observations. Importantly, we note that the results of Section 3.6.4.1 still hold: we are more likely to measure a MW-like kinematic plane in the presence of an LMC analogue near first pericenter relative to the general simulation sample. These caveats are detailed further in Appendix B.1.

As a more rigorous test, we examine the instances of planarity for which simulations are simultaneously spatially thin and kinematically coherent. We do not find any such simultaneously thin and coherent instances during $z = 0 - 0.2$ in the simulations. However, looking further back in time to $z = 0.5$, we find 10 snapshots that are simultaneously as thin and kinematically coherent as the MW satellites are today. This amounts to 0.3 per cent of the total sample of snapshots over $z = 0 - 0.5$. This level of simultaneous spatial and kinematic planarity agrees with Pawlowski & Kroupa (2020), who find that thin and coherent MW-like planes occur in < 0.1 per cent of IllustrisTNG hosts, than when we examine individual plane metrics. Notably, the instances of simultaneous planarity in our simulations occur in 2 out of 14 hosts (m12b and m12z). In both cases, the simultaneous spatial and kinematic planarity occurs around the time of the first pericentric passage of a massive ($M_* \geq 10^8\,\mathrm{M}_\odot$) satellite galaxy. The massive satellite that passes near m12b meets our criteria for being an LMC analogue. We explore the influence of LMC-like companions further in Section 3.6.4.1.

We do not see a significant difference in planarity between satellites of isolated hosts and satellites of hosts in LG-like pairs. Both the medians and ranges of plane metrics for each host type are essentially the same, so we do not further separate our results by host type. In Section 3.5.3, when we compare true satellites distributions to statistically isotropic distributions, the paired and isolated hosts do not appear systematically different from each

other either. This is consistent with results from Pawlowski et al. (2019), who reported no significant differences in planarity between dark matter-only simulations of isolated MW-mass halos and paired LG-like halos in the ELVIS simulations (Garrison-Kimmel et al., 2014).

3.6.1.2. *M31-like planes.* For comparison to M31's satellites, we mimic the completeness of PAndAS in our simulations. We first select all simulated satellites within $d_{\rm host} \leq 300\,{\rm kpc}$. Then, we randomly choose a line of sight from which to observe the simulation, and we select only the satellites that fall within a (2D) projected radius of 150 kpc from the host galaxy. We choose the 15 satellites with greatest stellar mass that fall within our mock PAndAS-like projection, to match the number of M31 satellites in our observational sample. We repeat this process along 10^3 random lines of sight.

In order to meet the 15 satellite criteria, we do not impose a lower limit on the stellar mass of satellites. At $z = 0$, the lowest mass satellite included in this sample has $M_* \approx 1.8 \times 10^4\,{\rm M}_\odot$. While most simulations easily meet the 15 satellite criteria, there are a few hosts with snapshots that have fewer than 15 luminous satellites within the mock survey area, so we exclude these snapshots. For example, at $z = 0$, four of the isolated hosts have fewer than 15 satellites selected (as few as 9 satellites) for some lines of sight. All simulations meet the satellite quota along most lines of sight, and in particular the hosts in LG-like pairs never suffer from this issue. The results that we achieve with this satellite number selection method are essentially the same as for a stellar mass selection method ($M_* \geq 10^5\,{\rm M}_\odot$). We use the full 3D phase space coordinates of these satellites to calculate spatial plane metrics, because the 3D spatial coordinates of each satellite within the coverage of PAndAS are well known. We calculate planarity metrics along each of 10^3 lines of sight at each snapshot over $z = 0 - 0.2$ for each simulated host.

Figure 3.3 shows that when considering the 15 most massive satellites, M31-like planes are common in our simulations. In particular, the axis ratios of simulated satellite systems

are typically as planar as the full sample of M31 satellites. More than 10 per cent of simulations are more planar than M31 for RMS height, so M31 is slightly thinner than our average simulation, but still lies within $\sim 1\sigma$ of the simulation median. Furthermore, throughout $z = 0 - 0.2$ the simulations have many instances of satellite configurations that are simultaneously as spatially thin and kinematically coherent as M31's satellites are under our selection criteria.

Radial (line-of-sight) velocities are currently the only kinematic information available for all of M31's satellites that we consider, so we cannot compute the 3D orbital pole dispersion of them as we did for the MW's satellites. We quantify kinematic coherence of satellites using $f_{v_{\rm los}}^{\rm max}$, where a larger fraction indicates greater kinematic coherence (see Section 3.5.2 for details). As Figure 3.3 shows, 14 per cent of simulations are more kinematically coherent than M31's satellites, though this is still within about 1σ of the simulation median. None of our simulations have all satellites sharing the same sense of orbital direction. Buck et al. (2016) have pointed out that a 2D metric like $f_{v_{\rm los}}^{\rm max}$ likely overestimates the true 3D kinematic coherence, so we may be overestimating the kinematic coherence in both our simulated and observed samples. The velocity coherence plot (right panel) is shown as a histogram because the underlying distribution is essentially discretely binned. Because each satellite population contains 15 satellites, the fraction of satellites sharing coherent velocities varies from 0.53 to 1.0 in steps of ~ 0.07 (see Section 3.5 for calculation details).

We find that the M31 satellite population as a whole is not significantly more planar than our simulations. This agrees with Conn et al. (2013) who found that M31's overall satellite population is consistent with a statistically isotropic distribution of satellites, though the 15 most-planar of its satellites lie within an exceptionally thin (12 kpc) plane. While Buck et al. (2015) use a different plane fitting method different from ours (a fixed-height plane), they also recover many instances of satellite planes as thin as the most-planar subset of M31 satellites.

We stress that our comparison to observations is not predicated on selecting the most planar subset of satellites in either simulations or observations. This is because we prioritize a wholistic view of the planarity of the satellite population as a whole, rather than highly planar subsets of those satellites. Other than having coherent LOS velocities, which do not unambiguously indicate orbital coherence, the member satellites of M31's plane are not significantly different from non-members, suggesting that they do not have different formation mechanisms or evolutionary histories (Collins et al., 2015). In addition, sampling the satellite distributions to calculate plane metrics is computationally expensive (see Section 3.4), and this is made more difficult by finding optimal planes for all satellite combinations. We defer such an investigation to future work.

For the rest of this work, we do not investigate M31-like planes further. Instead, we examine MW-like planes, given that completeness is more certain out to the virial radius, and precise 3D velocities of MW satellites are available. The availability of 3D velocities of MW satellites provides a more realistic metric of kinematic coherence.

3.6.2. Statistical significance and lifetimes of planes.

3.6.2.1. *Statistical significance of planes.* We now move from absolute metrics of planarity to a more general investigation of planarity, that does not rely on MW or M31 observations to establish what constitutes a planar configuration. We characterize the statistical significance of satellite planes in our simulations by randomizing the positions and velocities of satellites in order to form a statistically isotropic distribution as a control sample (see Section 3.5.3 for how we set this up). By generating 10^4 isotropic iterations and acquiring plane metrics from them, we create a bank of plane metrics that one might expect to measure if the distribution is statistically isotropic. This isotropic bank is used to compute plane significance by calculating the fraction ($f_{\rm iso}$) of isotropic iterations that are *more* planar than the true measured value at each snapshot. In effect, this provides an estimate of the probability of finding a thinner or more coherent plane in a random distribution of satellites.

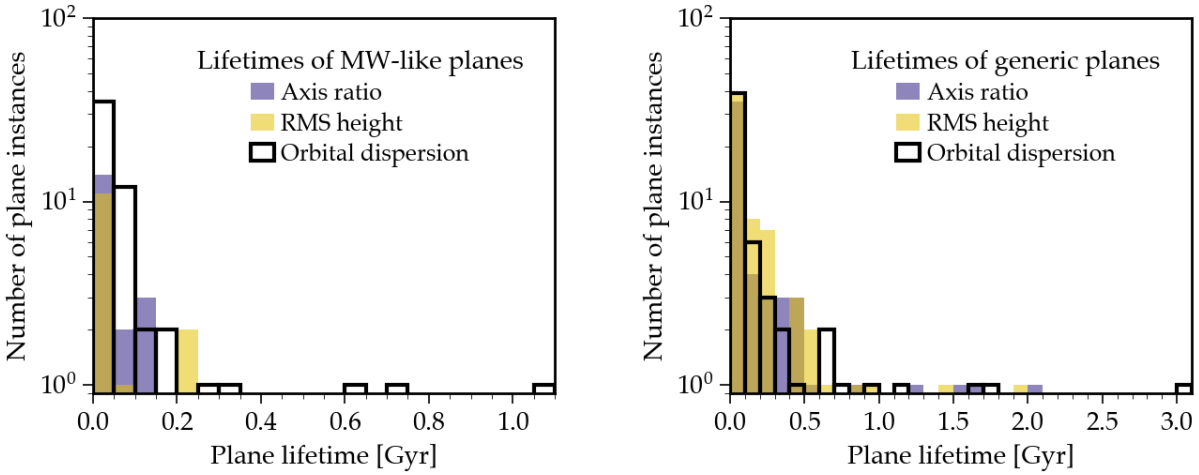

FIGURE 3.4. Satellite plane lifetimes measured over $z = 0-0.5$ (219 snapshots per host, ~ 25 Myr spacing) for the 14 satellites with the greatest M_* within $d_{\rm host} \leq 300\,{\rm kpc}$. We define lifetimes independently for each plane metric. *Left:* MW-like planes are those that have plane metrics at or below the MW upper one sigma limits. We have applied the same completeness correction for seeing through the disc as in Section 3.6.1.1. Such planes are rare and short-lived, with most lasting < 0.5 Gyr and none surviving for longer than $\sim$1 Gyr. Two out of the three instances of MW-like planes lasting >500 Myr occur in hosts that experience a pericenter passage of an LMC-like satellite. *Right:* Generic planes are any flattened or kinematically coherent systems whose plane metrics fall below the lower 68 percent limits of our simulations shown in red in Figure 3.6. Generic planes are also typically short-lived and many last for only a single snapshot. Half of the hosts have an instance of a generic plane that last > 1 Gyr, and two of those experience an LMC-like pericenter passage. While some generic planes live for a few Gyr, those planes are not typically simultaneously spatially thin and kinematically coherent.

Small fractions ($f_{\rm iso} \leq 0.05$) indicate a rare plane with high significance, while larger fractions ($f_{\rm iso} \geq 0.5$) show that the measured plane is consistent with an isotropic distribution of satellites.

We distinguish between two different measures of plane statistical significance: conditional probability and marginalized probability (following Cautun et al. (2015)). Marginalized probability refers to the significance of a system's planarity relative to an ensemble of planarity measurements on that system where the number of satellites considered is allowed to vary from the minimum number of points needed to define a plane (3) to some

maximum.We concentrate our analysis on conditional probability, because it represents the significance of a system's planarity given a certain set of constraints (such as completeness or total number of satellites). We calculate the significance of planes on simulations across $z = 0 - 0.2$, and on the observed positions and velocities of MW satellites. Again, for the simulations, we remove satellites obscured by the host disc at $|b_c| \leq 12°$. This is the same selection that we used in Figure 3.2.

By these simple metrics, and without correcting for selection or the look-elsewhere effect, the MW's plane is highly significant relative to a statistically isotropic distribution. Less than one per cent of the MW's isotropic realizations of its satellites have a thinner plane ($f_{\rm iso} = 0.003$ for RMS height or axis ratio), or a more kinematically coherent plane ($f_{\rm iso} = 0.005$ for orbital pole dispersion). In comparison, many of our simulation snapshots have median $f_{\rm iso} \gtrsim 0.5$, indicating that they are broadly consistent with and have no meaningful degree of planarity relative to a statistically isotropic distribution of satellites. See Appendix B.2 for a visual representation of $f_{\rm iso}$ for each host during $z = 0 - 0.2$. About half of both the isolated and paired hosts have median $f_{\rm iso} < 0.5$, and this similarity indicates that the paired host environment does not significantly enhance the statistical significance of satellite planes. About half of the hosts have $\sim 5 - 10$ per cent of their snapshots with $f_{\rm iso} < 0.05$, indicating significant spatial planes for these particular snapshots.

Only 3 out of the 14 hosts have significant kinematic coherence relative to a statistically isotropic distribution of satellite velocities, consistent with previous studies (e.g. Metz et al., 2008; Pawlowski & McGaugh, 2014; Ahmed et al., 2017; Pawlowski & Kroupa, 2020). Notably, none of the hosts have satellites that are simultaneously highly spatially significant ($f_{\rm iso} < 0.05$) and highly kinematically significant relative to a statistically isotropic distribution at any snapshot during $z = 0 - 0.2$. In general, hosts that with small (< 0.25) median $f_{\rm iso}$ for spatial planarity metrics do not have correspondingly small $f_{\rm iso}$ for kinematic coherence (orbital pole dispersion), and vice versa. While our simulations contain instances

of planes that are simultaneously as spatially thin and kinematically coherent at the MW in an absolute sense (by directly comparing plane metrics), the planes found in our simulations are not as significant relative to a statistically isotropic distribution.

3.6.2.2. *Lifetimes of planes.* Thus far we have focused our analysis on the spatial and kinematic coherence of satellite galaxies in our simulations over $z = 0 - 0.2$ ($\sim$2.4 Gyr). In this section, we seek to understand if the satellite planes we find are long-lived and stable, or merely transient configurations, across $z = 0 - 0.5$ ($\sim$ 5.1 Gyr, 219 snapshots). This longer time baseline allows us to examine the time evolution of satellite plane structures as satellites make multiple orbits around their host. A satellite in the inner regions of its host's halo may complete an orbit in under 1 Gyr. Satellites in the most outer regions of the host halo take $\sim$3-4 Gyr to undergo a complete orbit. We consider a plane to be "long-lived" if it persists for $\geq$ 1 Gyr, lasting for at least one satellite orbital timescale in the inner halo. We deem any planar configurations lasting $<$ 1 Gyr to be "short-lived", and we consider those lasting $<$ 500 Myr to be "transient" alignments that do not indicate coherence amongst satellite orbits because they are so short.

We examine the distribution of plane lifetimes over $z = 0 - 0.5$ separately for MW-like planes and generically flattened satellite systems. We define MW-like planes as those with plane metric values at or below the upper 68 per cent limits on MW values: RMS height $\leq$ 28 kpc, axis ratio $\leq$ 0.24, or orbital pole dispersion $\leq 67^{\circ}$. We measure MW-like plane lifetimes on the same simulation data in Figure 3.2, which includes a correction for seeing through the host disc. 'Generically' flattened means having plane metric values: RMS height $\leq$ 48 kpc, axis ratio $\leq$ 0.39, or orbital pole dispersion $\leq 71^{\circ}$, defined by the lower 68 per cent limit on simulation plane metrics during $z = 0 - 0.2$. We measure generic planes on the simulation data presented in Figure 3.6, which selects the 14 most massive satellites in stellar mass but does not include a correction for seeing through the host disc. We measure plane lifetimes ($\Delta t_{\rm plane}$) as the amount of time that a system spends consecutively at or below these plane

metric thresholds. Whether a satellite system is planar for only a single snapshot ($\lesssim$ 25 Myr) or many consecutive snapshots, we count it as a single instance of planarity.

Figure 3.4 shows that for both MW-like and generic planes, most planar instances are transient alignments and many last for just one snapshot ($\Delta t_{\rm plane} < 25\,{\rm Myr}$). There are 348 snapshots with MW-like planes in our simulations across all hosts over $z = 0 - 0.5$ (219 snapshots per host, 3,066 in total) and amongst all three 3D plane metrics. Out of the total 89 separate instances of MW-like planes, most (56) are in kinematic coherence and only one of them lasts for $\gtrsim$ 1Gyr. This only occurs for one host, m12b, which also happens to experience a close passage of an LMC-like satellite during that time, that we discuss further in Section 3.6.4.1. There are 1,796 snapshots and 177 separate instances of generically flattened planes in our simulations. Almost all generic planes last < 1 Gyr, with only a small fraction (< 10 per cent) of separate instances extending up to 3 Gyr. One host, m12f, has a generic kinematic plane lasting 3 Gyr and also experiences an LMC-like passage during this time. We conclude that satellite planes in our simulations, regardless of exact plane metric, are typically transient alignments that do not indicate a long-lived orbiting satellite structure, though the presence of LMC-like satellites can lead to longer-lived planes.

We also examine our simulations for instances of satellite configurations that are simultaneously spatially thin and kinematically coherent. We use the same plane metric thresholds as above to look at how often a satellite system meets the kinematic threshold and at least one of the spatial thresholds at the same snapshot. We do not find any instances of simultaneously thin and coherent MW-like planes over $z = 0 - 0.2$ using either our fiducial selection method ($N_{\rm sat} = 14$ and $d_{\rm host} \leq 300\,{\rm kpc}$) or combining that with a completeness correction due to seeing through the host's galactic disc. However, there are several instances of coincident thinness and coherence over $z = 0 - 0.5$, especially when we apply a completeness correction for seeing through the host disc. In particular, m12b and m12r have up to 13 snapshots ($\sim$ 325 Myr) of simultaneous spatial and kinematic planarity over $z = 0 - 0.5$.

We also consider the coincidence of generic planes, and find that m12b, m12r, and m12f all have snapshots with simultaneous spatial and kinematic planes even without implementing a completeness correction for the host disc. Both m12f and m12b have LMC satellite analogues during this time, as we discuss in Section 3.6.4.1. Interestingly, none of our hosts in LG-like pairs exhibit simultaneous planarity, reinforcing the result that LG-like host environments are not more likely to have satellite planes.

Shao et al. (2019) looked at plane lifetimes in the EAGLE simulations. They considered both a different sample size ($N_{\rm sat} = 11$) and a longer time baseline ($z \approx 0 - 2 \approx 10.5$ Gyr). This leads them to identify thinner planes in an absolute sense, because fewer satellites create a thinner plane. This time window may also catch some MW-like hosts as they are still being formed by mergers of smaller galaxies and before they have been able to form most of their stellar mass (e.g. Santistevan et al., 2020). However, they too found that most instances of MW-like spatially thin planes were short-lived (< 1 Gyr), but some systems remain orbitally coherent for upwards of 4 Gyr. Though we do not find such long-lived kinematic planes in our sample, this generally agrees with our findings.

Fernando et al. (2017, 2018) examined the stability of M31-like planes in idealized simulations, and found that most planes are short-lived and plane stability is highly sensitive to initial satellite phase space coordinates, plane alignment with the host halo, and subhalo abundance. The authors found that satellites moving perpendicular to the plane, misalignment of the plane with the halo axes, and increased subhalo abundance all generally caused planes to disrupt within ≤ 3 Gyr. While they demonstrated this within idealized simulations and specifically for comparison to the M31 plane, their modeling approach was general enough to compare to our plane lifetime results, where we find similarly short lifetimes for generic planes in cosmological simulations. This might lead one to conclude that the MW's plane is short-lived. However, we note that two of the three instances of MW-like planes with longer lifetimes in our simulations (based on orbital pole dispersion) have something else in

common, the presence of an LMC-like satellite. Such a massive satellite near pericenter that has brought with it its own satellites may contribute to a longer plane lifetime, so the MW's plane may not be as short-lived as the majority of our simulated MW-like planes. This is discussed further in Section 3.6.4.1.

3.6.3. Selection effects on measured planarity.

3.6.3.1. *Observational incompleteness from the host disc.* In our analysis of MW-like planarity thus far, we have applied a fixed obscuration correction for seeing through the host disc, masking out everything that lies within $|b_c| \leq 12^\circ$ (where c indicates a galactocentric coordinate system). We now analyze how the relative incidence of MW-like planes changes as a function of how much of the sky is obscured by the host's disc. We vary the region obscured by the galactic discs of simulated hosts from $b_c = 0^\circ$ (completely unobscured) to $|b_c| \leq 45^\circ$ (majority obscured) in increments of $\Delta|b_c| = 3^\circ$. For each obscured region we select the 14 most massive satellites in M_* within $d_{\rm host} \leq 300\,{\rm kpc}$ of a host to use in the plane sample. We define the relative incidence of MW-like planes as follows: we compute the fraction of snapshots with MW-like planes for each obscured region, and normalize it to the unobscured ($|b_c| = 0^\circ$) fraction of snapshots with MW-like planes. We repeat this process for each plane metric individually. However, for $|b_c| \geq 30^\circ$ there are typically fewer than 14 luminous satellites in the unobscured region and near $|b_c| \sim 40^\circ$ there are only about 10 satellites available on average, so we cannot draw strong conclusions about completeness effects in those limits.

Figure 3.5 shows the incidence of MW-like planes, measured independently for each metric, as a function of disc obscuration angle. We find that such incompleteness artificially boosts the fraction of snapshots with MW-like spatial planes for any value of $|b_c| > 0$. In particular, near the fiducial obscuration we adopt for MW-like planes in previous sections ($|b_c| = 12^\circ$), the incidence of MW-like planes is increased by about an order of magnitude. For $|b_c| \lesssim 40^\circ$, disc obscuration has a much smaller and opposite effect on kinematic planarity

compared to spatial planarity; MW-like kinematic planes tend to be somewhat washed out by incompleteness. Near our fiducial obscuration for the MW, the relative incidence of MW kinematic planes is about 0.77. As expected, disc obscuration has the largest effect on planarity when $|b_c| \sim 45°$, where so much of the sky is obscured that any detected satellites would appear to be in a plane purely due to incompleteness.

Our results show that observational incompleteness from the host disc can have a strong effect on measured spatial planarity. If the MW's satellite population is incomplete from seeing through the Galactic disc at our fiducial level, then MW observations may be overestimating the spatial planarity of MW satellites by a factor of $\sim 10 - 20$. To a much lesser degree, MW observations may underestimate the kinematic coherence of satellites by a factor of $\sim 1 - 2$. Because this incompleteness may bias our analysis of the underlying causes of satellite planes, we only use a host disc correction when comparing directly to MW observations in Sections 3.6.1.1 – 3.6.2.2 and 3.6.4.1. For the remainder of this paper, we do not include a host disc correction.

3.6.3.2. *Method of selecting simulated satellites.* We also explore how using a fixed number selection for satellites compared to using a stellar mass threshold affects planarity measurements. Our primary method of satellite selection throughout this work is to choose the 14 most massive satellites by rank-ordering them in stellar mass, because the number of satellites in a sample strongly correlates with the measured planarity (e.g. Pawlowski et al., 2019). In terms of observational completeness and resolution in simulations, another way to select satellites may be to impose a simple stellar mass threshold. So we test our fixed number selection against a stellar mass threshold method: $M_* \geq 10^5\,\mathrm{M}_\odot$ and $d_{\rm host} \leq 300\,\mathrm{kpc}$. However, this leads to a range of numbers of satellites selected around each host, which makes it difficult to compare plane metrics across simulations and observations. The total number of satellites with $M_* \geq 10^5\,\mathrm{M}_\odot$ and $d_{\rm host} \leq 300\,\mathrm{kpc}$ per host varies from 10-31 during $z = 0 - 0.2$ in our simulations.

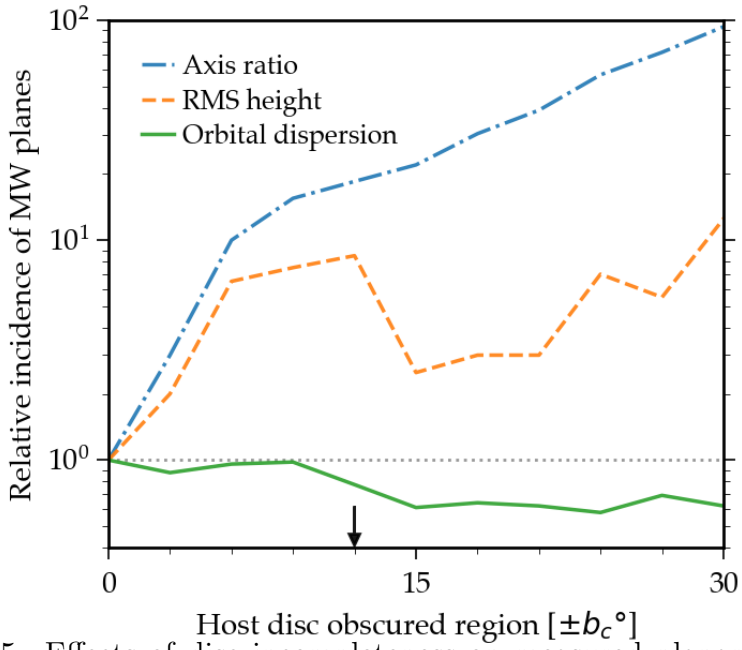

FIGURE 3.5. Effects of disc incompleteness on measured planarity. We define the relative incidence of MW planes as the fraction of snapshots during $z = 0 - 0.2$ with MW-like planes normalized to the true or unobscured fraction ($|b_c| = 0°$). We select the 14 most massive satellites in M_* within $d_{\rm host} \leq 300\,{\rm kpc}$ of each host, but for $|b_c| \geq 30°$ there are usually fewer than 14 luminous satellites available. The horizontal line represents consistency with the unobscured fraction. The arrow shows the fiducial obscuration that we adopt for MW-like planes, $|b_c| = 12°$. Spatial planarity ($c/a \leq 0.24$, $\Delta_{\rm h} \leq 28$ kpc) is much more affected by host disc obscuration than kinematic planarity. Spatial planarity jumps an order of magnitude between $|b_c| = 0°$ and $|b_c| \sim 10°$. Kinematic planarity ($\Delta_{\rm orb} \leq 67°$) is slightly diminished by host disc obscuration. At $|b_c| = 12°$, we are $8.5 - 18.5\times$ more likely to measure a MW-like spatial plane and $1.3\times$ less likely to measure a MW-like kinematic plane. As expected, when nearly half of the sky is obscured spatial planarity is highly likely to be measured.

Figure 3.6 shows that planes with $N_{\rm sat} = 14$ tend to be both thinner and more kinematically coherent than planes with $M_* \geq 10^5\,{\rm M}_\odot$, because while some $M_* \geq 10^5\,{\rm M}_\odot$ satellite populations have $N_{\rm sat} < 14$, more actually have $N_{\rm sat} > 14$. One consequence is that when using the M_* selection the simulations never reach the MW's RMS height (27 kpc) during $z = 0 - 0.2$, but the fixed number selection does. The small bump in the $N_{\rm sat} = 14$ orbital pole dispersion distribution is from a single host, m12f, during the snapshots following a

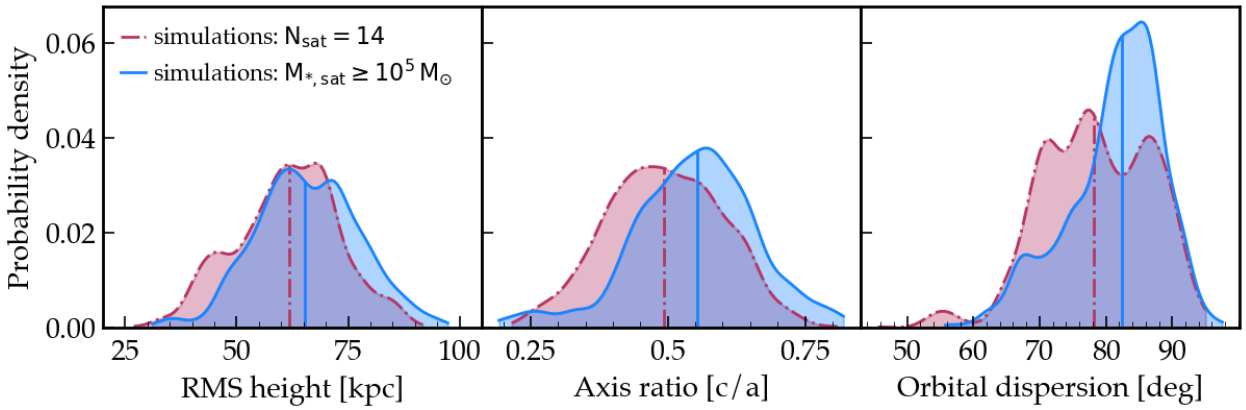

FIGURE 3.6. Planarity of simulated satellite galaxies ($d_{\rm host} \leq 300\,{\rm kpc}$) selected using a fixed number method versus a stellar mass threshold. Note that we do not include a correction for completeness due to seeing through the host galactic disc here. We generate KDEs using 114 snapshots over $z = 0 - 0.2$ for each of the 14 simulated hosts. The vertical lines are the medians for each distribution. The red distributions are the 14 most massive satellites in stellar mass, while the blue distributions are all satellite galaxies with $M_* \geq 10^5\,{\rm M}_\odot$. Thin and coherent planes are rare in the simulations using these particular selections and time baseline, but using the number selection for satellites yields lower (more planar) metrics because the stellar mass selection allows for many more satellites to be included ($N_{\rm sat} = 10 - 31$).

close passage of an LMC-like satellite. We discuss effects of such an LMC-like companion further in the following section.

This selection exercise highlights an important aspect of the satellite plane problem: many of the conclusions drawn about the nature of satellite planes are sensitive to satellite selection method, likely because of underlying sensitivity to the number of satellites in the sample. Had we used the stellar mass threshold as our fiducial selection method in previous sections, we would have found more evidence for tension between simulations and observations, but deciding whether that tension is cosmologically significant is hampered by the sensitivity of plane metrics to both incompleteness and sample selection.

3.6.4. Exploring physical explanations of planes.

3.6.4.1. *Influence of an LMC-like satellite.* The presence of a massive satellite galaxy near pericenter, like the Large Magellanic Cloud (LMC), has been suggested as a possible

explanation for the dynamical origin of the MW's satellite plane (Li & Helmi, 2008; D'Onghia & Lake, 2008), from the accretion of multiple satellites in a group with the LMC (e.g., Wetzel et al., 2015b; Deason et al., 2015; Jethwa et al., 2016; Sales et al., 2017; Jahn et al., 2019). We seek to determine whether or not the presence of an LMC-like companion has an effect on the planarity in simulations. We compare planarity metrics measured on systems experiencing an LMC-like passage to those without an LMC-like passage. We identify pericentric passages of four LMC-mass analogues in our simulations based on the following selection criteria:

(i) $t_{\rm peri} > 7.5$ Gyr ($z < 0.7$)
(ii) $\rm M_{sub,peak} > 4 \times 10^{10}\,M_\odot$ and $M_* > 10^9\,\rm M_\odot$
(iii) $d_{\rm peri} < 50$ kpc
(iv) The satellite is at its first pericentric passage.

This broad time window allows us to capture a larger number LMC-like passages, which tend to be rare as we have defined them. The minimum mass is consistent with measurements of the LMC's mass (Saha et al., 2010), and the maximum pericenter distance reflects the measured distance and orbit of the LMC (Freedman et al., 2001; Besla et al., 2007; Kallivayalil et al., 2013). Table 3.1 lists the four hosts in our simulations with LMC satellite analogues that meet these criteria, all of which are from simulations of isolated MW-like hosts rather than paired/LG-like hosts. We emphasize that these satellites are not the only sufficiently massive satellites in the simulations, but that they are the only instances that satisfy all our LMC analogue criteria simultaneously.

To compare planarity during LMC-like passages and otherwise, we first select all snapshots within ± 5 snapshots (a time window of ~ 250 Myr) of the LMC-like pericenter passage in each of the four simulations containing an LMC analogue. This gives us a total of 44 snapshots that we classify as occurring close enough to an LMC analogue pericenter to exhibit any dynamical effects of group infall. We compare plane metrics from those snapshots to

TABLE 3.1. Properties of the LMC satellite analogues at their first pericentric passage about their MW/M31-mass host in our FIRE-2 simulations. We select satellites with $M_{sub,peak} > 4 \times 10^{10}\,M_\odot$ and $M_* > 10^9\,M_\odot$ that have their first pericenter after 7.5 Gyr ($z < 0.7$) and within 50 kpc of their host. t_{peri} is given in Gyr and d_{peri} is given in kpc.

Host	$M_{sub,bound}$ $[10^{10}\,M_\odot]$	$M_{sub,peak}$ $[10^{11}\,M_\odot]$	M_* $[10^9\,M_\odot]$	t_{peri}	z_{peri}	d_{peri}
m12b	12.0	2.1	7.1	8.8	0.49	38
m12c	5.1	1.6	1.2	12.9	0.07	18
m12f	6.0	1.5	2.6	10.8	0.26	36
m12w	4.9	0.8	1.3	8.0	0.59	8

plane metrics measured on all other simulations (excluding the four hosts with LMC analogues) up to the earliest snapshot included in the LMC sample ($z \sim 0 - 0.7$, 247 snapshots per host). We apply our fiducial disc obscuration correction, masking out all satellites within $|b_c| \leq 12^\circ$ of the hosts' galactic discs in our simulations. To calculate plane metrics we select $N_{sat} = 14$ of the most massive satellites ranked by stellar mass.

Figure 3.7 summarizes our results for the planarity of satellites during an LMC analogue pericenter passage compared to all other satellite systems during $z \sim 0 - 0.7$. In general, the presence of an LMC analogue leads to thinner and more kinematically coherent satellite planes on average. The presence of an LMC analogue shrinks the range of spatial plane metric values and slightly shifts them towards smaller (thinner) values. In particular, the range of axis ratios is much smaller in the presence of an LMC analogue. The right panel of Figure 3.7 also shows that the presence of an LMC analogue is correlated with more of the simulation distribution having tighter orbital alignment of satellites. For all three metrics, we are $\sim 2 - 3$ times more likely to measure a MW-like plane during an LMC pericentric passage compared to the general simulation sample. This result persists if we widen our time window to ± 10 snapshots (~ 500 Myr). The enhancement in the fraction of snapshots with MW-like spatial planes and an LMC near pericenter washes out for time windows larger than ~ 500 Myr, but the enhancement for MW-like kinematic planes persists in even the

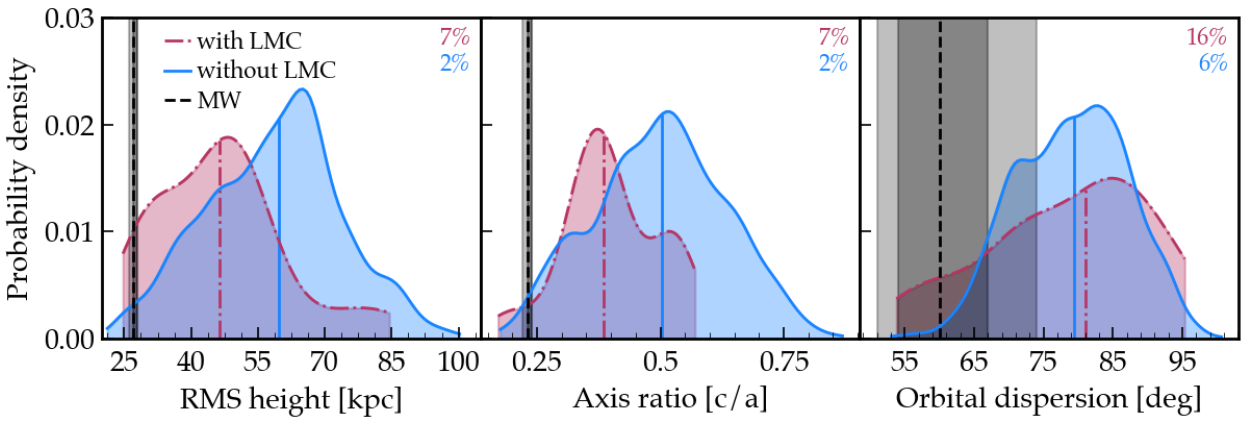

FIGURE 3.7. Planarity of satellites of hosts experiencing a first pericenter passage of an LMC satellite analogue (red) compared to all other hosts without an LMC-like passage (blue). We rank order satellites by stellar mass and choose the 14 most massive around each host within $d_{\rm host} \leq 300\,{\rm kpc}$. We select snapshots within ± 125 Myr of LMC-like passages that occur during $z \sim 0 - 0.7$. Only 4 hosts have such LMC-like passages (see Table 3.1). Vertical colored lines are the medians of the simulation distributions. MW planarity values are the vertical black lines and shaded regions. We apply a disc obscuration correction and omit satellites within $|b_c| \leq 12°$. LMC passages push towards ~ 20 per cent lower plane metric medians and smaller ranges of spatial planarity metrics. MW-like planes are $\sim 2 - 3$ times more likely to be measured during an LMC-like passage.

largest time window (± 40 snapshots or ~ 2 Gyr) that we tested. If we instead use the 'best-available' proper motion data set from Pawlowski & Kroupa (2020), the increased likelihood of measuring a MW-like kinematic plane is actually strengthened, because only 1 per cent of the general snapshot sample during $z \sim 0 - 0.7$ is MW-like versus 11 per cent near LMC analogue pericenters (see Appendix B.1). Thus, the presence of the LMC on first infall may contribute significantly to the thin and (even more so) kinematically coherent satellite plane around the MW.

We also consider the time evolution of planarity both before and after LMC analogue pericentric passages. For RMS height and orbital pole dispersion in particular, the hosts begin to experience downward trends in these metrics just before or at the time that the LMC analogue crosses within $R_{\rm 200m}$ of the MW-mass host, reaching minimum values up to

few hundred Myr after the LMC analogue's pericentric passage. As an additional comparison of simulated and observed MW satellite kinematics, we calculate the velocity anisotropy parameter, β, following Cautun & Frenk (2017). We measure $\beta = -1.35 \pm 0.2$ for our sample of 14 observed MW satellites, indicating a preference for circular orbits. The distribution of β for the simulations experiencing LMC analogue pericenter passages has a longer and more prominent tail towards more negative values of β (more circular orbits), as well as a lower median value than the general simulation sample. Thus, our analysis of β also suggests that the presence of the LMC may increase the likelihood of measuring MW-like satellite kinematics. While more circular orbits could conceivably lead to a more stable satellite plane, Cautun & Frenk (2017) found little evidence for a correlation between aligned orbital poles and circularity of satellite orbits. We leave a full dynamical analysis of the LMC's influence on planarity for future work.

We find that the main reason for enhanced planarity in systems with LMC analogues is that the LMC analogues bring satellites with them that are counted in the plane sample, and because it is only at first pericenter there has not been enough time for the LMC and its satellites to dissociate from each other (e.g., Deason et al., 2015). The four LMC analogues each bring in $2-4$ satellites with $M_* \geq 10^5\,\mathrm{M}_\odot$, consistent with the results presented in Jahn et al. (2019) for both likely satellites of the LMC and FIRE-2 simulation predictions for satellites of LMC-mass hosts. Of the $2-4$ LMC analogue satellites, $1-3$ of them are counted toward the $N_{\rm sat} = 14$ satellites in the plane sample. The host with the most planar configuration that we find (m12b), which also has instances of simultaneous spatial and kinematic planarity, brings in four satellites with $M_* \geq 10^5\,\mathrm{M}_\odot$ and three of these (plus the LMC analogue itself) are counted in the plane calculations. This means that the LMC analogue and its satellites account for ~ 30 per cent of the plane sample for m12b, so spatial and kinematic coherence of the LMC subgroup can easily drive the measured plane metrics to lower values.

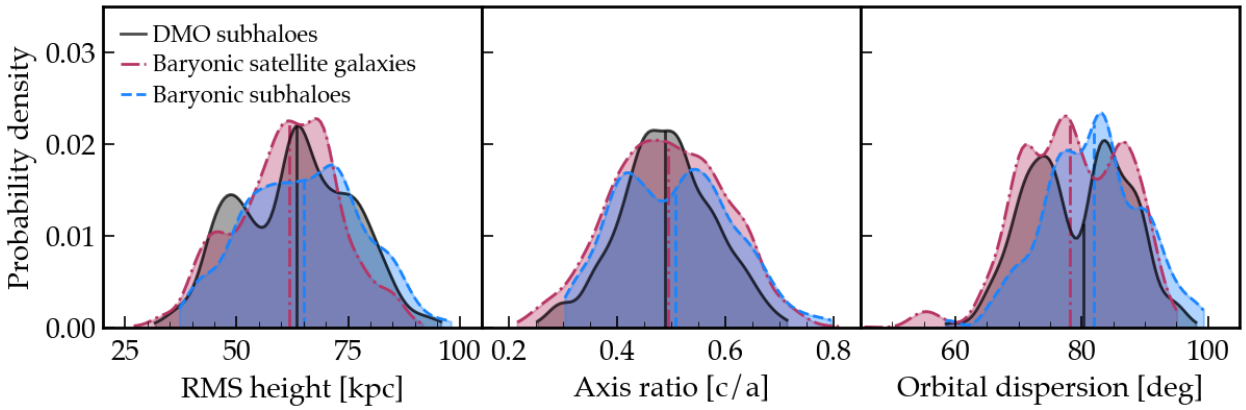

FIGURE 3.8. Planarity in baryonic versus dark matter-only (DMO) simulations. We compare the 14 most massive baryonic satellites and subhaloes to DMO subhaloes within $d_{\rm host} \leq 300\,{\rm kpc}$. We rank order subhaloes by $M_{\rm peak}$ and satellites by M_*. We generate KDEs using 114 snapshots per host over $z = 0 - 0.2$ for each of the hosts and the vertical colored lines are the medians of each distribution. All three samples show similar planarity, but a few systems have baryonic satellites with greater kinematic coherence. We conclude that there are no significant differences in planarity of baryonic versus DMO simulations.

Shao et al. (2018) examined whether anisotropic accretion or group accretion could explain the formation of satellite planes. They ultimately found that most massive satellites were singly accreted, and that anisotropic accretion rather than group accretion correlated more with planarity. In light of this we test for whether planarity correlates in general with group accretion and average infall times. We find that overall most of the satellites in our sample were either singly accreted or accreted as groups of two. We do not find a strong correlation between such group accretion and planarity in our full sample. So we stress that our key result is that only a sufficiently massive LMC-like satellite near first pericenter shows a clear sign of enhancing planarity. Two hosts with LMC analogues (m12f and m12b) experience extended periods of planarity ranging from $0.7 - 3$ Gyr (Section 3.6.2.2). We conclude that LMC-like satellites contribute significantly to satellite planes with moderate lifetimes ($\sim 1 - 2$ Gyr), but that they are unlikely to have a permanent effect on the satellite distribution on longer timescales.

3.6.4.2. *Baryonic versus dark matter-only simulations.* We also ran all of our simulations without baryonic physics, except for one of the isolated hosts, m12z. We compare planarity of these dark matter-only (DMO) simulations to our baryonic simulations in order to investigate potential baryonic effects on satellite planes, given that many previous studies of planes have used DMO simulations. We ran the DMO simulations with the same number of DM particles and the same gravitational force softening. We compare planes in our baryonic simulations to planes in their DMO counterparts by selecting luminous satellites and dark matter subhaloes within $d_{\rm host} \leq 300\,{\rm kpc}$. We choose the 14 most massive object from each sample by rank ordering satellite galaxies by M_* and subhaloes using $M_{\rm peak}$.

Figure 3.8 shows the distributions of planarity for satellite galaxies and subhaloes, both selecting the top 14 subhaloes by $M_{\rm peak}$ and the top satellite galaxies by M_*, which are not identical samples because of scatter in the $M_* - M_{\rm peak}$ relation. The satellite galaxy distributions (red) are identical to those in Figure 3.6. While the three distributions in each panel have slightly different shapes, they have almost the same ranges and medians. Using a rank ordering selection, the planarity of DMO subhaloes is essentially identical to that of baryonic satellites. We find that this general result is robust with respect to rank ordering subhaloes by different properties such as $M_{\rm halo}$, $V_{\rm peak}$, and $V_{\rm circ}$. The one exception is a small population of baryonic satellites that extend to lower orbital dispersion values during the passage of an LMC-like satellite (see Section 3.6.4.1).

These results are surprising in light of the differences in the radial distributions of satellites and subhaloes in our simulations, wherein DMO subhaloes are more radially concentrated around their host than luminous satellites (Samuel et al., 2020). One might expect to find thinner planes in DMO simulations because subhaloes reside spatially closer to the host halo. Alternatively, one might expect baryonic simulations to show greater planarity, given that the surviving population is biased to more tangential orbits (e.g. Garrison-Kimmel et al., 2017b). Ahmed et al. (2017) observed both a difference in the significance and satellite

TABLE 3.2. Correlations between planarity in simulations and properties of the host and radial distribution of satellites. We select satellite galaxies by rank ordering them by stellar mass at each snapshot and choosing the 14 most massive. We measure host halo properties using only dark matter. We quote the correlation coefficients (r) and p-values given by the Spearman correlation test. For brevity, we only show correlations with $p < 0.1$, though we note that only $p \lesssim 0.01$ indicates a significant correlation in our sample.

Planarity metric	Host/system property	r	p-value
RMS height	Host halo concentration	0.49	0.07
	Host M_*	0.55	0.04
	Host halo axis ratio (c/a)	0.60	0.02
	Host M_*/M_{halo}	0.54	0.04
	R_{50}	0.68	0.01
Axis ratio	Host halo axis ratio (c/a)	0.52	0.06
Orbital dispersion	R_{90}/R_{10}	0.59	0.03
	R_{90}/R_{50}	0.58	0.03

membership of their planes in baryonic versus DMO simulations of the same four systems. Satellite membership here refers to whether the satellites contributing to planes belong to the same subhaloes in baryonic and DMO runs of the same systems. While we do not explicitly consider differences in satellite membership, we do find that our DMO spatial planes are also typically more significant relative to a statistically isotropic distribution of satellites than their baryonic counterparts, with most having $P < 0.5$ during $z = 0 - 0.2$. However, the significance of DMO kinematic planes is on par with the baryonic simulations. So while we find that the significance of spatial planes may be slightly overestimated in DMO simulations relative to baryonic simulations, the absolute planarity is not much different from that in baryonic simulations. If, instead, we select subhaloes at a fixed value of $M_{\rm peak} \geq 8 \times 10^8\,{\rm M}_\odot$, DMO simulations typically have many more subhaloes meeting this criteria. This difference in number of subhaloes in the plane sample reduces planarity in DMO simulations, because planes with more members are generally less planar (Pawlowski et al., 2019).

3.6.4.3. *Correlations between plane metrics and host-satellite system properties.* Finally, we explore relationships between satellite planarity and host and satellite system properties,

but we find few correlations. We quantify correlation using the Spearman correlation coefficient (r) and p-value, applied to the median value of each plane metric and host property over $z = 0 - 0.2$ for all hosts. None of the correlations that we found are particularly strong, as all have $r < 0.7$. We summarize correlations in Table 3.2, where we only show correlations with $p < 0.1$ for brevity. Only correlations with $p \lesssim 0.01$ indicate a statistically significant correlation in our sample, and there is only one correlation meeting this criteria.

We considered four host properties: stellar mass, dark matter halo mass ($M_{200\mathrm{m}}$), stellar-to-total mass ratio, and halo concentration. Both of the spatial metrics correlate with the host halo axis ratio, such that more triaxial host halos are more likely to have thinner satellite planes. RMS height is also correlated with host stellar mass, whereby more massive host discs may act to disrupt rather than promote thin planes. Orbital dispersion does not correlate significantly with any of the host properties. While there is some evidence for spatial planarity correlating with host halo axis ratio, the correlations are not strong ($0.5 < r < 0.6$).

We conclude that it is unlikely that host properties drive the formation of satellite planes because we do not find strong and consistent correlations between planarity and host properties. We also explored the alignment of planes with respect to both the host galaxy disc and the host halo minor axis, but we found no conclusive correlations among our sample. Given the polar satellite plane around the MW, and the M31 satellite plane being more aligned with the host disc, our results support that we expect no consistent correlation with the disc. This results agrees with Pawlowski et al. (2019), who found that satellite plane metrics did not correlate with host properties like halo concentration or halo formation time in dark matter only simulations.

We also test for correlations between planarity and the radial distribution of satellites. The strongest correlation that we find ($r = 0.68$ and $p = 0.01$) exists between RMS height and R_{50}, the radius enclosing 50 per cent of the satellites. This correlation may arise from

more satellites being near pericenter, rather than actually being flattened into a thin plane, because RMS height is a dimensional quantity (unlike dimensionless axis ratio), so we would expect it to correlate with satellite distances. We also examine planarity as a function of R_{90}/R_{50} and R_{90}/R_{10}, where R_{90}/R_{50} is the ratio of the distance from the host that encloses 90 per cent of the satellite population to the distance from the host that encloses 50 per cent of the satellite population, and R_{90}/R_{10} is similarly defined. These ratios describe the radial concentration of the satellites around their host, and they are the only metrics that significantly correlate with orbital dispersion. In both cases, more concentrated satellite systems are correlated with less kinematically coherent planes.

We investigated relationships between planarity and properties of major mergers in the histories of the host galaxies. We adopt the following definition of major merger: a merger occurring during $z = 0 - 3$ with a stellar mass ratio of at least 10 per cent. Altogether, 10 of the 14 hosts experience at least one major merger. m12c, m12f, m12r, m12z, Louise, and Remus each have one major merger, while m12m, Thelma, and Romulus each have two, and m12b has a total of three. Six of the hosts experience mergers that we broadly classify as similar to the Gaia-Enceladus event (Belokurov et al., 2018; Helmi et al., 2018) by requiring them to have occurred between $\sim 8 - 11$ Gyr ago and to have a stellar mass ratio of $10 - 30$ per cent. Two of the hosts (m12m and Thelma) each experience two Gaia-Enceladus type mergers. Three out of the four hosts with LMC analogues (m12b, m12c, and m12f) experience at least one major merger, and m12b and m12c each experience a merger within the Gaia-Enceladus time window, but with mass ratios (13 and 35 per cent, respectively) just outside of our nominal range. We tested for correlations between planarity during $z = 0 - 0.2$ and the number of major mergers per host, the timing of the last major merger, and the mass ratio of the last major merger, but found no significant correlations. The strongest correlations were between orbital pole dispersion and merger properties, but the maximum correlation coefficient was only ~ 0.4 with a p-value of ~ 0.2, indicating overall

weak correlations between planarity and major mergers in the host. While there is perhaps some evidence for a correlation between LMC analogues (arguably the most important factor in creating satellite planes) and major mergers, we conclude that past major mergers do not have a strong independent influence on planarity.

3.7. Summary and Discussion

We explored the incidence and origin of planes of satellite galaxies in the FIRE-2 simulations, using satellites around 14 MW/M31-mass galaxies over $z = 0 - 0.2$. We compared to and provided context for satellite planes in the Local Group, including all satellites with $M_* \geq 10^5\,\mathrm{M}_\odot$ around the MW and within the PAndAS survey of M31. We summarize our main results as follows.

3.7.1. Rareness of Planes.

- MW-like planes exist in our simulations, but they are relatively rare among our randomly selected $\sim 10^{12}\,M_\odot$ halos at $z = 0 - 0.2$: planes at least as thin or coherent as the MW's plane occur in $\sim 1-5$ per cent of all snapshots, and planes as thin and coherent according to *spatial and kinematic metrics simultaneously* occur in ~ 0.3 per cent of snapshots.
- *However*, if we select halos that feature a LMC-mass satellite analogue near its first pericentric passage, then the frequency of MW-like or thinner planes dramatically increases to $7-16$ per cent, with $\sim 5\%$ at least as thin as the MW plane by spatial and kinematic metrics simultaneously.
- If we consider M31's satellite population as a whole, the planarity of satellites around M31 is common in our simulations. By every spatial or kinematic (or simultaneous) measure we consider, M31's satellites lie within $\sim 1\sigma$ of the median of randomly selected halos of similar mass that we simulated.

- Most of our simulations are not significantly planar relative to a statistically isotropic distribution of satellites.

3.7.2. Physical Origins of Planes.

- Most MW-like thin satellite planes are transient and last $< 500\,\mathrm{Myr}$ in our simulations. However, the presence of an LMC satellite analogue near pericenter produces longer lifetimes of $\sim 0.7 - 3\,\mathrm{Gyr}$. More generically flattened satellite systems survive for up to $\sim 2 - 3\,\mathrm{Gyr}$, even without requiring a massive satellite like an LMC analogue.
- We do not find significant differences in planarity of satellites around hosts in Local Group-like pairs versus isolated hosts.
- Dark matter-only (DMO) simulations show no significant differences in planarity compared to their baryonic-simulation counterparts, when selecting a fixed number of satellites in each sample.
- Correlations between plane thickness and other satellite population properties (radial concentration) or host properties (mass, concentration, size, axis ratio) are generally modest or weak. Plane thickness is generally larger for more radially extended satellite distributions, as expected. The one property that strongly correlates with the presence of spatially thin and kinematically coherent planes is the presence of an LMC analogue near first pericentric passage.

3.7.3. Observational and Selection Effects.

- Plane metrics can be sensitive to the satellite selection method in simulations and observations. Selecting just the 14 satellite galaxies with the highest stellar mass in the simulation produces thinner planes compared to selecting *all* satellites with $M_* > 10^5\,M_\odot$, because the latter tends to select more satellites, which produces thicker planes.

- Incompleteness from the inability to see through the host galaxy disc (as in the MW) can increase the probability of measuring MW-like spatial planes by as much as a factor ~ 10. This bias is opposite in sign but much smaller for kinematic planes.
- We have *not* corrected in any of our analysis for any 'look-elsewhere' effects, including the choice to look for 'planes', the choice of definition of 'plane', sample selection, number of satellites, etc. These corrections only would *decrease* the statistical significance of the observed planes, as outliers from simulations.

3.7.4. Discussion. Though only $1-2$ per cent of snapshots for all 14 hosts during $z = 0 - 0.2$ contain satellite planes at least as thin as the MW's, we do not interpret this as a strong tension with ΛCDM cosmology. Instead, we identify the mere presence of MW-like planes in the simulations as evidence that cosmological simulation indeed can form thin planes of satellites, as long as they have adequate mass and spatial resolution. We find that planes are much more common in the presence of LMC analogues, as suggested by Li & Helmi (2008) and D'Onghia & Lake (2008), which provides evidence that future work should prioritize comparing the MW against simulations with an LMC analogue. Considering the entire M31 satellite population, M31-like satellite planes are common in our simulations, and combined with the fact that our simulations are only marginally more planar than a statistically isotropic distribution of satellites, this may indicate that M31's satellites as a whole are not significantly planar. Our most promising result points to the presence of the LMC near first pericenter as a likely primary driver of planarity. The lack of strong correlations between planarity and other properties of the host-satellite systems leaves us with few other physical explanations for the MW's highly coherent satellite plane. If our simulations are representative of the MW, then the observed MW plane is likely to be a temporary effect that will wash out in subsequent orbits of the LMC (Deason et al., 2015).

Pawlowski & Kroupa (2020) report similar percentages ($\sim 2-3$ per cent) of hosts with MW-like planes in the IllustrisTNG simulations, but contrary to our own conclusion, they

claim that this does in fact constitute a challenge to ΛCDM. Comparing the percentages in Pawlowski & Kroupa (2020) to those that we obtain in this study is a readily understandable synthesis of the two studies, but we now highlight a few key differences in the underlying data sets (both observed and simulated) and analysis that warrant a more nuanced comparison of our work with that of Pawlowski & Kroupa (2020). Our observational sample includes three additional MW satellites (Crater II, Antlia II, and Canes Venatici I) that meet our nominal stellar mass criteria ($M_* \geq 10^5\,\mathrm{M}_\odot$). Pawlowski & Kroupa (2020) may have excluded these satellites because they are borderline cases of 'classical' dwarfs given their diffuse morphology and/or stellar masses, or perhaps because some were discovered only recently and thus not ideal for the historical comparison in that work.

In calculating orbital pole dispersion in simulated systems, Pawlowski & Kroupa (2020) sample from each host's satellites to select the most aligned subsample, something that we do not explore here. The use of a 'most-aligned' sub-sample could lead us to measure smaller orbital pole dispersions in our systems (and hence a higher fraction of snapshots with MW-like kinematic planes), but we are limited in sample size given our stellar mass criteria. The percentages from Pawlowski & Kroupa (2020) represent robust statistical significances, because their data are comprised of over 1000 independent host systems, whereas our data include multiple snapshots for each of only 14 independent host systems. In comparison, our measured percentages are not strict statistical significances.

Furthermore, our claim that Local Group satellite planes are not a strong challenge to ΛCDM rests on two main conclusions from our work that are not in Pawlowski & Kroupa (2020): the presence of an LMC analogue makes measuring a MW-like plane more likely (and indicates a system that is a better match to the MW), as well as the commonality of M31-like planes. We conclude that, because MW-like planes are less rare when we match the MW's satellite population more precisely than in previous studies of satellite planes, and because we only compare to one observed system in this case, that observed satellite planes

do not constitute a strong challenge to ΛCDM. The rareness of satellite planes remains an interesting topic, but we maintain that a *strong* challenge to ΛCDM cosmology requires *strong* evidence of rarity (as opposed to mere uncertainty), which we do not find in this work.

We have deliberately approached our analysis of satellite planes as agnostically as we can. In choosing a fixed number of satellites for our nominal selection method, we have tried to both show the clearest comparisons between our simulations and LG observations, as well as mitigate the confounding effects of correlations between $N_{\rm sat}$ and planarity. Further studies of the most-planar subsamples of simulated satellites, as examined in Pawlowski et al. (2013) and extended in Santos-Santos et al. (2020), may yield more insight into the nature of satellite planes. We defer an analysis of satellite sub-samples to future work.

We also have not yet considered a comparison to satellite systems outside of the LG. There is evidence for satellite planes outside of the LG around Centaurus A (Müller et al., 2018), and recent studies have examined planarity around hosts in SDSS (Ibata et al., 2014a; Brainerd & Samuels, 2020) and the SAGA survey (Mao et al., 2021). Connecting LG hosts to a statistical sample of similar hosts will be crucial in evaluating the significance of planar alignments and the validity of proposed formation mechanisms, demonstrating the need for large surveys with e.g., the Nancy Grace Roman Space Telescope, which promises to significantly augment the observational sample of MW analogues. LG galaxies are also aligned with large scale structure, along a local sheet, which is not captured in our simulations and may play a part in the formation of satellite planes (Neuzil et al., 2020). Simulations that can accurately reproduce this large scale structure may offer new insight into satellite planes (Libeskind et al., 2020).

Acknowledgements

We thank Marcel Pawlowski for insightful comments and discussion that improved this manuscript.

This research made use of Astropy,[3] a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013, 2018), the IPython package (Pérez & Granger, 2007), NumPy (Harris et al., 2020), SciPy (Jones et al., 2001), Numba (Lam et al., 2015), and matplotlib, a Python library for publication quality graphics (Hunter, 2007).

JS, AW, and SC received support from NASA through ATP grants 80NSSC18K1097 and 80NSSC20K0513; HST grants GO-14734, AR-15057, AR-15809, and GO-15902 from the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555; a Scialog Award from the Heising-Simons Foundation; and a Hellman Fellowship. We performed this work in part at the Aspen Center for Physics, supported by NSF grant PHY-1607611, and at the KITP, supported NSF grant PHY-1748958. PFH was provided by an Alfred P. Sloan Research Fellowship, NSF grant #1715847 and CAREER grant #1455342, and NASA grants NNX15AT06G, JPL 1589742, 17-ATP17-0214. MBK acknowledges support from NSF CAREER award AST-1752913, NSF grant AST-1910346, NASA grant NNX17AG29G, and HST-AR-15006, HST-AR-15809, HST-GO-15658, HST-GO-15901, and HST-GO-15902 from STScI. 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 simulations using the Extreme Science and Engineering Discovery Environment (XSEDE) supported by NSF grant ACI-1548562, Blue Waters via allocation PRAC NSF.1713353 supported by the NSF, and NASA's HEC Program through the NAS Division at Ames Research Center.

[3]http://www.astropy.org

CHAPTER 4

Satellite quenching

4.1. Abstract

The star formation and gas content of satellite dwarf galaxies in the Local Group (LG) appear to be highly sensitive to the environment around the Milky Way (MW) and Andromeda (M31). We use 240 satellite dwarf galaxies from the FIRE-2 simulations to examine environmental trends in quenching around 14 MW/M31-mass ($M_{200m} \approx 1 - 2 \times 10^{12}\,M_\odot$) hosts. All low-mass ($M_* = 10^{5-7}$) satellites are quenched at $z = 0$, and the quenched fraction of satellites decreases with increasing satellite stellar mass, similar to the LG. However, the quenched fractions are inconsistent with satellites from the SAGA survey, especially at low satellite stellar mass. We measure higher average quenched fractions around more massive hosts, indicating a response in quenching to the details of the host environment. There is a strong positive correlation between when quenching occurs relative to first infall into the host halo and the stellar mass of a satellite, which supports the idea that the host environment dominates the quenching process. However, we also find evidence for multiple pathways to quenching for satellites before and after entering the host environment, including ram pressure stripping, group preprocessing, and internal stellar feedback.

4.2. Introduction

Dwarf galaxies are especially susceptible to both their own internal feedback and environmental influence on their evolution because their shallow gravitational potentials are insufficient to retain perturbed gas. At the low-mass end ($M_* \lesssim 10^5\,M_\odot$), ultrafaint dwarfs are likely quenched by reionization, unable to shield their gas from the dissociative effects of

UV photons in the early Universe (Bullock et al., 2000; Weisz et al., 2014; Rodriguez Wimberly et al., 2019). At intermediate dwarf masses ($M_* \sim 10^{6-8}\mathrm{M}_\odot$), the cold gas that fuels star formation is routinely heated and ejected from dwarf galaxies by their bursty episodes of star formation and the ensuing feedback from supernovae. For isolated dwarf galaxies, the gas ejected by feedback is often able to cool and re-accrete onto the dwarf for further star formation (Geha et al., 2012), but satellite dwarf galaxies are subjected to additional disruptive forces in the host galaxy's environment. For example, the relative motion of satellites through the hot halo gas of the host environment causes satellites to experience ram pressure, which can act in concert with feedback to more efficiently remove rarefied gas from satellites and quench their star formation. Thus, satellite dwarf galaxies are more likely to be quenched than their isolated counterparts.

Local Group (LG) satellite dwarf galaxies are almost all gas-poor and no longer forming stars, except for some of the most massive satellites: the Magellanic Clouds, LGS3, and IC10 (Wetzel et al., 2015b). However, results from the Satellites Around Galactic Analogs (SAGA) survey indicate that nearly all satellites of nearby isolated MW analogues are still star-forming at $z \sim 0$ (Geha et al., 2017; Mao et al., 2021). Since the host galaxies in SAGA are chosen specifically to match the MW in stellar mass, and because the number of satellites down to their completeness limit are in broad agreement with the LG satellite population and simulations (Samuel et al., 2020), this difference in the quenched fraction of satellites presents a unique tension. Putman et al. (2021) recently published a compilation of updated neutral hydrogen (HI) detections in LG dwarfs, revealing that these dwarfs are increasingly HI-poor as distance from the closest massive galaxy (either the MW or M31) decreases, and also as distance from the surface of the LG (which bounds the virial radii of both the MW and M31) decreases. This may imply that the paired nature of the MW and M31 may have an additional disruptive effect on satellite galaxies compared to an isolated

MW-like environment, possibly due to a more massive or dense gaseous halo encompassing both hosts.

In addition to environmental quenching, satellites of MW-mass hosts at $z = 0$ may actually become quenched before entering the host galaxy's environment through various processes such as reionization, internal stellar feedback, and group preprocessing. Reionization is thought to be the primary quenching mechanism for low-mass dwarfs, especially in the ultra-faint ($M_* \sim 10^{3-4}\,\mathrm{M}_\odot$) regime (Weisz et al., 2014). Stellar feedback within dwarf galaxies may also be an efficient quenching mechanism because gas is removed from their shallow gravitational potentials by bursts in their star formation histories followed by supernovae explosions that drive outflows. Furthermore, the same environmental effects that act to quench satellite dwarf galaxies in a MW-mass host's halo can also manifest in lower-mass groups (Wetzel et al., 2015a), such as the recently demonstrated ram pressure stripping of dwarfs in the vicinity of simulated LMC-mass hosts (Jahn et al., 2021). These preprocessing effects, like ram pressure stripping, dynamical friction, and tidal disruption, could bolster the perceived environmental effects of a MW-mass host galaxy on satellites that were previously part of a group, so it is important to understand different quenching mechanisms in order to disentangle them from effects of the host environment.

4.3. Simulations

4.3.1. FIRE simulation suite. We use the Latte and ELVIS on FIRE cosmological hydrodynamic simulation suites from the Feedback In Realistic Environments (FIRE) project[1] (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a,b). The baryonic mass resolution varies between the different simulations in the range $\mathrm{m_{baryon,ini}} = 3500-7100\,\mathrm{M}_\odot$. The gas softening is fully adaptive, matched to the hydrodynamic resolution, and the minimum gas resolution (inter-element spacing) and softening length reached in Latte is ≈ 1 pc. The halos chosen

[1] `https://fire.northwestern.edu/`

for zoom-in simulation have halo masses $M_{200m} = 1 - 2 \times 10^{12}\,M_\odot$, within observational uncertainties of the MW's halo mass[2]. The simulations we use in this work reproduce the mass functions, radial distributions, and star formation histories of classical ($M_* \geq 10^5\,M_\odot$) dwarf galaxies around MW/M31-like hosts (Wetzel et al., 2016; Garrison-Kimmel et al., 2019a,b; Samuel et al., 2020).

We ran all simulations with the FIRE-2 implementations of fluid dynamics, star formation, and stellar feedback (Hopkins et al., 2018). FIRE uses a Lagrangian meshless finite-mass (MFM) hydrodynamics code, GIZMO (Hopkins, 2015). GIZMO enables adaptive hydrodynamic gas particle smoothing depending on the density of particles while still conserving mass, energy, and momentum to machine accuracy. Gravitational forces are solved using an upgraded version of the N-body GADGET-3 Tree-PM solver (Springel, 2005). The FIRE-2 methodology includes detailed subgrid models for gas physics, star formation, and stellar feedback. Gas models used include: a metallicity-dependent treatment of radiative heating and cooling over $10 - 10^{10}$ K (Hopkins et al., 2018), a cosmic ultraviolet background with early HI reionization ($z_{\rm reion} \sim 10$) (Faucher-Giguère et al., 2009), and turbulent metal diffusion (Hopkins, 2016; Su et al., 2017; Escala et al., 2018). We model several stellar feedback processes including core-collapse and Type Ia supernovae, continuous stellar mass loss, photoionization, photoelectric heating, and radiation pressure.

For all simulations, we generate cosmological zoom-in initial conditions at $z = 99$ using the MUSIC code (Hahn & Abel, 2011). All simulations assume flat ΛCDM cosmologies, with slightly different parameters across the full suite: $h = 0.68 - 0.71$, $\Omega_\Lambda = 0.69 - 0.734$, $\Omega_m = 0.266 - 0.31$, $\Omega_b = 0.0455 - 0.048$, $\sigma_8 = 0.801 - 0.82$, and $n_s = 0.961 - 0.97$, broadly consistent with Planck Collaboration et al. (2018).

4.3.2. Halo finder. Using the ROCKSTAR 6D halo finder (Behroozi et al., 2013a), we identify dark matter (sub)halos in our simulations at each of the 600 snapshots per simulation

[2]'200m' indicates a measurement relative to 200 times the mean matter density of the Universe

and construct merger trees using CONSISTENT-TREES (Behroozi et al., 2013b). We assign star particles to dark matter halos in a post-processing step adapted from Necib et al. (2018) and further described in Samuel et al. (2020, 2021).

4.3.3. Assigning gas particles. We assigned gas particles to subhalos at $z = 0$ in a manner similar to the star particle assignment routine described in Samuel et al. (2020). First, we use a subhalo catalog to identify the center position, bulk velocity, and other properties of each subhalo. Then, in order to assign gas particles to subhalos, we require the gas particles to be (1) within twice the stellar half mass radius of a subhalo and (2) within twice the greater of the subhalo maximum circular velocity or the velocity dispersion of DM particles belonging to the subhalo. Each gas particle has a corresponding temperature, density, metallicity, and HI fraction, which we use to determine satellite gas properties.

4.4. Results

We select dwarf galaxies from our simulation volumes at $z = 0$ that have stellar masses $M_* \geq 10^5\,\mathrm{M}_\odot$, which we believe to reside in well-resolved dark matter halos (Samuel et al., 2020). We limit our study here to the 240 satellite dwarf galaxies within 300 kpc of their host, of which only 39 have any assigned gas particles. Gas-rich low-mass ($M_* < 10^7\mathrm{M}_\odot$) dwarfs only exist outside the host virial radius ($d_{\mathrm{host}} \gtrsim 250$ kpc) in the simulations, suggesting effective environmental gas removal in this mass regime. These results are consistent with observations of HI in LG dwarfs where only the most massive dwarfs within the LG have detectable HI (Spekkens et al., 2014; Putman et al., 2021).

4.4.1. Quenching timescales and modes. We define a galaxy as quenched if no stars have formed within it for the last 200 Myr or longer. We also consider an alternate definition of quenched based on the lack of any HI gas. Satellites of intermediate masses ($M_* \sim 10^{6-8}$) are often quenched by the host environment within 2 Gyr (Wetzel et al., 2015b; Fillingham et al., 2016; Rodriguez Wimberly et al., 2019), so we define a useful metric for each satellite

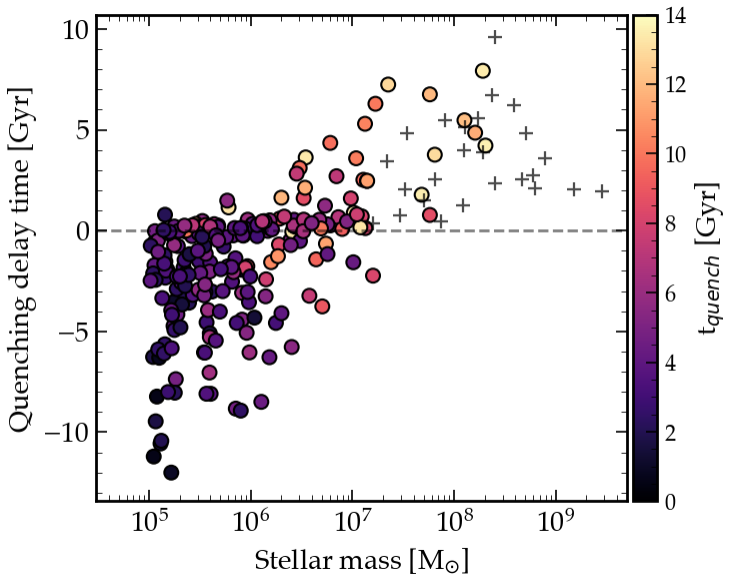

FIGURE 4.1. Quenching delay time (between infall and last star formation) versus stellar mass. Points are colored by time of last star formation, or are shown as crosses if still star-forming at $z = 0$. Low-mass satellites quenched earlier and either before or at infall, intermediate-mass satellites quench near infall with some scatter, and high-mass satellites either remain star-forming or quench after infall.

that we refer to as the quenching delay time: $t_{delay} = t_{infall} - t_{quench}$ that measures the time it takes for a satellite to quench after first infalling ($d_{\rm host} < {\rm R}_{\rm 200m,\,host}$) into the host halo.

Figure 4.1 shows t_{delay} versus stellar mass for all satellite galaxies in our simulations. Galaxies that remain star-forming at $z = 0$ are shown as unfilled symbols, and their quenching delay times are simply the lookback times to first infall for each satellite. More massive galaxies tend to quench after infall while less massive galaxies quench well before or during infall. There is a noticeable build-up of galaxies ($M_* = 10^{5-7}{\rm M}_\odot$) along the line of $t_{delay} = 0$, indicating that many satellites are quenched just as they arrive to the edge of the host's halo and that environmental effects are strongest in this mass regime. Quenching delay times extend to large negative values (and early cosmic times) in the smallest mass bin ($M_* = 10^{5-6}{\rm M}_\odot$) likely because these galaxies are fully or at least partially quenched by reionization (Fitts et al., 2017), however, the large spread in this bin (in both delay and

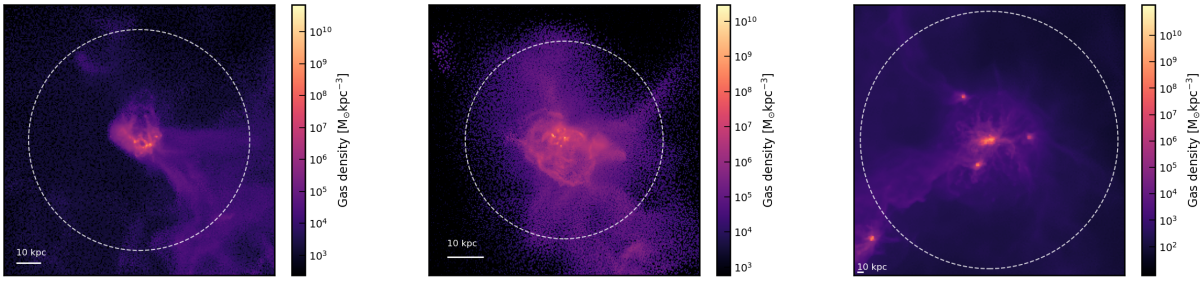

FIGURE 4.2. Images of satellite gas density at time of infall into host halo demonstrating the effects of internal feedback, group preprocessing, and interactions with the MW-mass host halo. From left to right: a satellite experiences intense ram pressure at infall and is devoid of gas by $z = 0$, stellar feedback rarefies the gas within a dwarf at infall, and a massive satellite is about fall into the MW-mass host halo with its own satellites.

cosmic time) also indicates a notably gradual transition from being quenched by reionization to quenched by infall.

Figure 4.2 shows examples of gas density maps for satellites at infall, hinting at a few possible modes of satellite quenching that likely act in concert to shut off star formation. The left panel is a classic example of ram pressure stripping as it flies through the host galaxy's ambient halo gas. This satellite was not massive enough to retain any of its gas and quickly ends up devoid of gas and is quenched by $z = 0$. The middle panel illustrates how even in more massive satellites, internal stellar feedback can rarefy and remove gas from the center of the galaxy, thereby potentially making it more susceptible to ram pressure effects in subsequent orbits through the host halo. The map in the right panel shows a small galaxy group (the central galaxy is a massive dwarf) that is just outside of the MW-mass host's halo. Prior group associations of satellites like these may act to preprocess their gas, further aiding in their environmental quenching within the MW-mass host halo (Jahn et al., 2021). Ram pressure stripping in a MW-mass halo is sufficient to quench most satellites at $M_* \lesssim 10^8\,\mathrm{M}_\odot$, as long as the CGM is sufficiently clumpy (Fillingham et al., 2016). However, in the absence

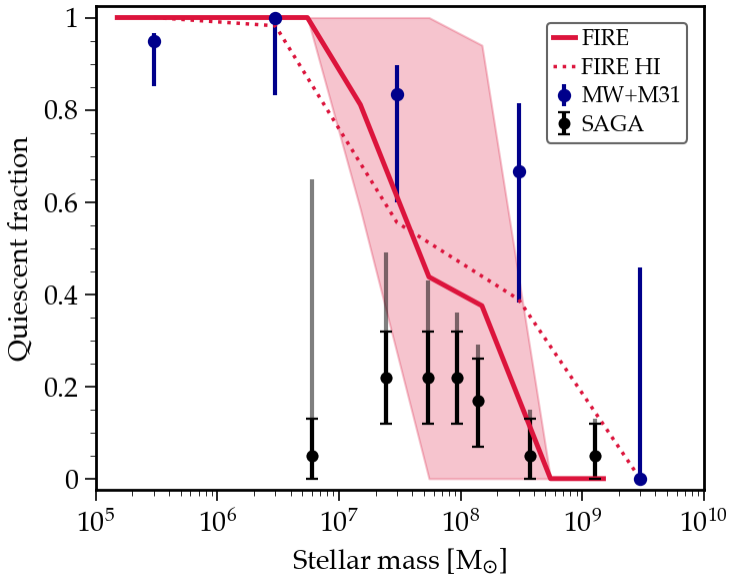

FIGURE 4.3. Quenched fraction versus stellar mass of all $z = 0$ satellites stacked together, compared to the Local Group and SAGA. The solid red line and shaded region show the median and 68% host-to-host scatter across the simulations. The Local Group error bars are statistical. The darker SAGA error bars are shot noise, while the lighter SAGA error bars are an estimate of the impact of maximal incompleteness. The simulations are in broad agreement with the higher quenched fractions of the Local Group, and SAGA satellites are more star-forming than both the simulations and the Local Group.

of dense clumps, preprocessing and perhaps internal stellar feedback are needed to first rarefy cold gas in satellites enough for it to be stripped in a MW-mass halo environment.

4.4.2. Quenched fraction. We now turn to the task of exploring the quenched fraction of satellites in an effort to understand the observed differences between LG and SAGA satellites. The SAGA hosts were chosen to be MW analogs ($-23 > M_K > -24.6$ or $M_* \approx 10^{10-11}\,\mathrm{M}_\odot$, consistent with the stellar masses of our simulated hosts, and thus should provide a statistical sample of observed hosts to compare against. Figure 4.3 shows the quenched fraction of satellites as a function of their stellar mass in simulations and observations. We generate the simulation results (solid red line and shaded region) by finding the fraction of satellites that have not formed stars in at least the last 200 Myr, and plotting the median and 68 per cent scatter across the 14 hosts in our sample. We show this as a function

of satellite stellar mass using half dex bins over $M_* = 10^{5-9.5}\,\mathrm{M}_\odot$. The results from the simulations are shown on top of the results from Wetzel et al. (2015b) for the LG (blue) and results from Mao et al. (2021) for the SAGA survey (black). Satellites in the LG are largely quenched, especially at $M_* < 10^7\,\mathrm{M}_\odot$. Though SAGA does not reach as low in stellar mass as observations of the LG, there is still a clear tension between the mostly star-forming SAGA satellites and the systematically higher quenched fraction in the LG. This difference persists even when considering robust estimates of SAGA's incompleteness; the darker error bars on the SAGA data are shot noise, while the lighter error bars are an estimate of the impact of maximal incompleteness, by making the reasonable assumption that the photometry of satellite candidates is complete to spectroscopic limits and the liberal assumption that all spectroscopic non-detections are quenched.

Note that each of the three data sets is binned differently by stellar mass, and uses a different definition of quenched. We define quenching in the simulations at the beginning of section 4.4.1, while the LG quenched definition is based on the absence of detectable HI, and SAGA's quenched definition relies on the absence of significant Hα emission. Our simulation results are qualitatively the same if we instead use an HI-deficient criterion (dotted red line) Because Hα probes star formation on short timescales ($\sim$ 5 Gyr, Flores Velázquez et al. (2021)), using a different quenching criterion for the SAGA data that probes longer timescales likely would not change their quenched fractions much, and could actually push to even lower (more discrepant) values. Similarly, taking into account extinction by gas and dust would also psuh SAGA quenched fractions to lower values.

Notably, the simulations are broadly consistent with the LG across the full range of satellite stellar masses ($M_* = 10^{5-9.5}\,\mathrm{M}_\odot$), and with SAGA in the intermediate to high mass range $M_* > 10^7\,\mathrm{M}_\odot$. At the low mass end ($M_* < 10^7\,\mathrm{M}_\odot$), both the simulations and the LG have a completely quenched population of satellites. At $M_* \sim 10^{8-9}\,\mathrm{M}_\odot$, the simulations appear to have a larger fraction of star-forming satellites than the LG, a region

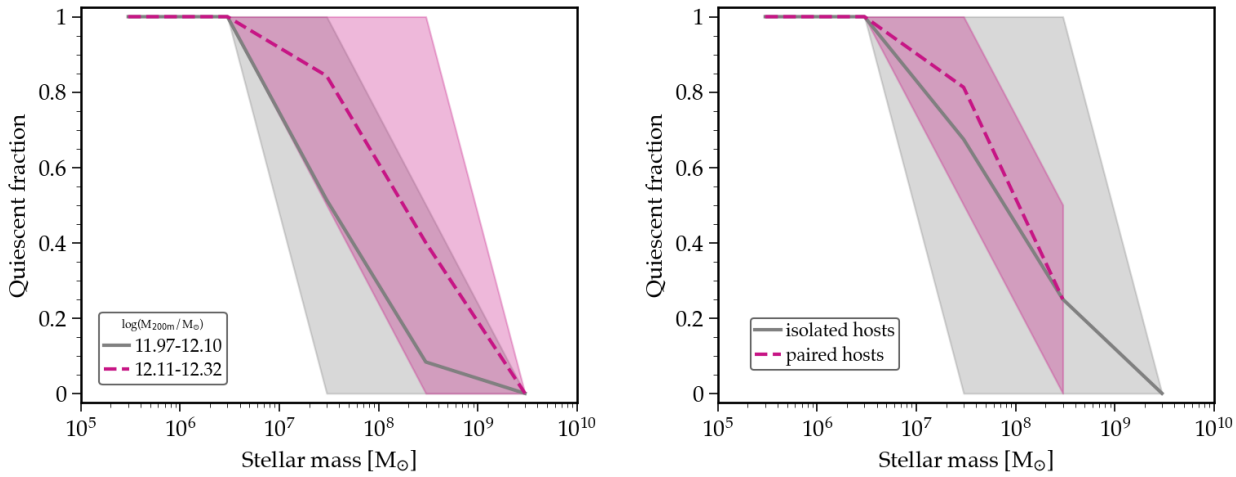

FIGURE 4.4. Quenched fractions of satellites versus stellar mass at $z = 0$. Scatter shows the full range across hosts, and black line is the median of that range. *Left:* Sample is split by host $M_{200\mathrm{m}}$. *Right:* Sample is split by paired or isolated environment. The more massive hosts tend to have a larger fraction of quenched satellites, but there is no significant difference between hosts in pairs and isolated hosts.

where the LG contains only a few satellites such as the SMC and M32, and where our simulation sample dwindles to only 18 satellites. The simulation and LG data converge again in the highest mass bin where all satellites are star-forming. The difference between the simulations and SAGA is only significant in our lowest mass bins ($M_* < 10^7\,\mathrm{M}_\odot$), where we expect observational incompleteness to be greatest. However, even with SAGA's liberal estimates of incompleteness at these masses, the differences between SAGA and either the LG or the simulations are still large. Similar and more severe tensions between simulation results and SAGA have been noted by other authors, but have yet to be explained (Akins et al., 2021; Karunakaran et al., 2021).

4.4.2.1. *Environmental trends.* We now attempt to explore some possible explanations for the differences in the quenched fractions in the LG and our simulations compared to quenched fractions in SAGA. We investigate the impact of the MW/M31-mass host environment by separating our sample in two ways. First we examine whether the total mass (including

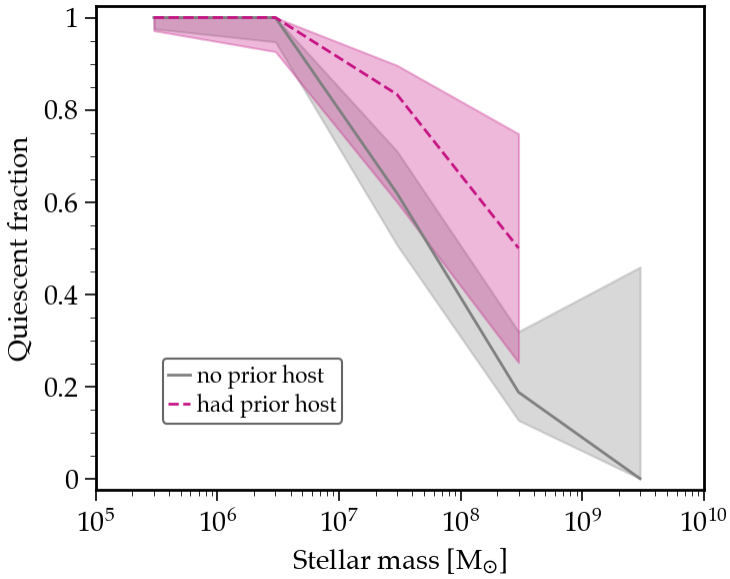

FIGURE 4.5. Quenched fractions of satellites versus stellar mass at $z = 0$, separated by whether or not a satellite was previously a member of a small group. Lines are the result of stacking all satellites across the different simulations together and the scatter is statistical. Satellites with $10^7\,M_\odot \lesssim M_* \lesssim 10^9\,M_\odot$ that are accreted in groups are more likely to be quenched than satellites of similar mass that were never part of another group.

stars, gas, and dark matter) within M_{200m} might have an effect on satellite quenching. We calculate a quenched fraction for each host and then divide our sample into low-mass hosts ($9.2 \times 10^{11}\,M_\odot < M_{200m} \leq 1.3 \times 10^{12}\,M_\odot$) and high-mass hosts ($1.3 \times 10^{12}\,M_\odot < M_{200m} < 2.1 \times 10^{12}\,M_\odot$). The left panel of Figure 4.4 shows the median and full (100%) scatter in quenched fractions across the hosts in each mass group. On average, more massive hosts tend to have higher quenched fractions and also scatter to higher values than less massive hosts, whereas less massive hosts tend to have lower median quenched fractions and scatter lower than more massive hosts. We repeat a similar exercise in the right panel of Figure 4.4, but this time separating hosts by whether they are a member of a LG-like paired system from the ELVIS on FIRE suite (6 hosts) or an isolated host from Latte. Here, we see that the paired hosts have a slightly higher median quenched fraction, but the scatter in paired hosts is completely encompassed by the scatter in isolated hosts, so these differences are not

significant. The paired hosts also lack more massive satellites to compare against in our highest mass bin.

These trends in host mass and group environment may hint at an underlying difference between SAGA hosts and the LG. If SAGA is somehow preferentially selecting less massive hosts (combined with their overt choice to select isolated MW hosts), then based on our results we might expect to see a smaller quenched fraction of satellites down to $M_* = 10^6\,\mathrm{M}_\odot$ similar to the upper limits on SAGA error bars in Figure 4.3. However, this would require us to assume that SAGA's incompleteness is at a level consistent with their worst case possible, and there may be further unexplored differences in host properties that could be responsible for the tension between SAGA and the LG/simulations.

4.4.2.2. *Group preprocessing.* In light of recent work by Jahn et al. (2021), who found signatures of environmental quenching of satellites around isolated LMC-mass galaxies, we now search for signs of pre-quenching in similar low-mass groups that fall into the MW-mass halos in our simulations. Using the merger tree of each simulation, we search for prior group associations for each satellite at $z = 0$. The merger trees record a DM halo as being a satellite of another halo if it passes within the host halo's radius, $\mathrm{R}_{\mathrm{200m}}$. If one of our $z = 0$ satellites was previously a satellite of another halo, we count it as being in a prior group association, regardless of if it was still a part of that group upon infall into the MW-mass host's halo or if its previous host survives as a satellite at $z = 0$. We end up with 94 of our $z = 0$ satellites that had prior hosts out of the total of 240. Note that we do not include the small group host/central galaxy in the count of satellites with prior group associations.

Figure 4.5 shows the resulting quenched fractions as a function of stellar mass in the simulations, separating satellites by whether or not they were previously hosted by a halo other than the main MW-mass host. We stack satellites across all simulations to produce the lines, and the scatter indicates expected statistical uncertainty assuming the fraction of quenched satellites in each bin is a beta continuous random variable. Similar to the satellites

with no prior host, all satellites with a prior host and $M_* = 10^{5-7}\,\mathrm{M}_\odot$ are quenched at $z = 0$. Our sample of dwarfs with prior hosts does not contain any galaxies with $M_* = 10^{9-10}$, so we cannot make a comparison in our highest stellar mass bin. Interestingly, at intermediate to high stellar masses ($M_* = 10^{7-9}$) the sample of satellites with prior hosts has a slightly higher median quenched fraction and scatter that reaches significantly higher than the sample without prior hosts. Because the intermediate mass range is where we expect to see the strongest signatures of environmental quenching, our result likely reflects the effects of group preprocessing either fully or partially quenching satellites before they enter the MW-mass host's environment. However, our result may also reflect the fact that satellites that were previously hosted by another galaxy are more likely to be found in denser environments, which may also have an effect on their gas content and quenching.

4.5. Conclusions

We have used the Latte and ELVIS on FIRE suites of 14 MW/M31-mass host galaxies and their satellites from the FIRE simulations to investigate the gas content and star formation quenching of satellite dwarf galaxies. Similar to the Local Group, we find that most of the simulated satellites are gas-poor and quenched at $z = 0$. When we examine the quenched fraction of satellites as a function of their stellar mass and stack all of our simulations together, we find that the FIRE simulations have quenched fractions generally consistent with satellites in the Local Group and SAGA. However, there is some slight tension with the highly star-forming dwarf satellites around MW analogs from the SAGA survey at the low end of satellite stellar masses we consider here ($M_* < 10^7\,\mathrm{M}_\odot$).

We have analyzed broad characteristics of the host environment that may be responsible for this difference such as total host mass and paired versus isolated host configurations. We find that both more massive hosts are more likely to have higher quenched fractions of satellites, especially for satellites with $M_* = 10^{7-9}\,\mathrm{M}_\odot$, but that the differences between isolated and paired hosts are not significant. We also find that satellites that were previously

hosted by another galaxy prior to falling into the MW-mass host environment are also more likely to be quenched compared to satellites without prior hosts. Our findings could indicate that SAGA hosts may have systematically smaller total masses than the MW/M31, but these results are not conclusive. However, a more nuanced (in progress) analysis of the gaseous halos around the simulated hosts is needed to determine more exact environmental trends and whether ram pressure stripping alone is responsible for quenching satellites, or if there are more internal processes are at play. We will also consider effects of different physics in simulations, namely the inclusion of cosmic rays and magnetohydrodynamics, on host CGM and satellite quenching in future work.

CHAPTER 5

Conclusion

I have used zoom-in cosmological hydrodynamic simulations of MW/M31-like galaxies and their satellite dwarf galaxies to provide new context for observations of the Local Group. With each project, I have examined the simulated satellites in more detail, helping to shift the narrative of this field from broad issues like the missing satellites problem to the more granular questions surrounding satellite dynamics and quenching. I summarize the results and implications of each chapter below.

Chapter 2 laid ground work for future studies with the Latte and ELVIS on FIRE simulations, establishing that they go beyond resolving the missing satellites problem and that they actually have radial distributions of satellite galaxies that are representative of the Local Group. I used this flexibility to make predictions for satellite discovery in the outer halos of both the MW and M31, finding that we should expect a few to several more dwarfs to be discovered around each host. Furthermore, the shapes of satellite radial profiles reveal that the MW's satellites are somewhat unusually concentrated, lending further evidence to the idea of potential incompleteness in the current census of MW satellites. I examined radial profiles of satellites in 2D projection to make comparisons to observations of satellites from the SAGA survey, finding broad agreement between simulations and satellite systems of MW analogs in the Local Universe. I also benchmarked radial profiles of luminous satellites in comparison to DMO subhalos, a useful tool for other studies using DMO simulations to compare to observations.

Chapter 3 tackled perhaps the most important outstanding small-scale challenge to structure formation, the satellite plane problem, showing that such satellite configurations are

possible, though rare, in simulations. The key result of this work was the identification of the LMC's possible role in shaping the MW's satellite plane, because thin and kinematically coherent satellite systems were significantly more likely to be measured in simulated systems experiencing a first pericenter passage of an LMC analogue. Additionally, the satellite plane of M31 appears to be less coherent and hence much more common in our simulations, at least in the absence of full 3D velocities for M31's satellites. I showed that most MW-like planes in the simulations are short-lived, which supports the idea that we are catching the MW's satellites at a unique time, likely during the LMC's first passage around the MW. Finally, my estimates of the effects of sample selection and observational incompleteness on satellite planes indicate that spatial coherence may be overestimated for the MW's satellites if the satellite census is incomplete because of the difficulty of detecting diffuse systems through the MW's disk.

Chapter 4 began a deeper look into the evolution of internal properties of satellite galaxies, namely their gas content and star formation quenching. The simulations display the same broad trends as observations of the Local Group at $z = 0$: low-mass satellites ($M_* \lesssim 10^7 \mathrm{M}_\odot$) are universally quenched (prior to or at infall) and devoid of gas, intermediate-mass satellites ($10^7 \mathrm{M}_\odot \lesssim M_* \lesssim 10^8 \mathrm{M}_\odot$) quench at various times before and after infall with a few retaining gas and forming stars, and high-mass satellites ($M_* \gtrsim 10^8 \mathrm{M}_\odot$) are mostly still star-forming. Considering the full sample of simulated satellites, the quenched fraction as a function of stellar mass is broadly consistent with the Local Group, and slightly in tension with the mostly star-forming satellites of MW analogs from the SAGA survey at the low-mass end. Satellites appear to experience various levels of ram pressure at infall, which removes gas and contributes to quenching their star formation. Differences in the total masses and circumgalactic medium of the Local Group and SAGA hosts may have something to do with their differing quenched fractions of satellites, but this needs to be investigated further.

Galaxy formation in the Local Group is a rapidly progressing field that has the potential to shed light on several areas of astrophysics in the near future. Determining the effect of a massive satellite like the LMC on both the MW and its satellites is just starting to become an attainable goal because of the increasing reach of stellar surveys like Gaia and others into the MW's stellar halo and because of better resolution in zoom-in simulations. However, the cost of resolution in simulations is a smaller sample of independent systems to compare to observations, so another major challenge is matching the statistical sample of MW analogs from SAGA and other surveys with a similar number of high resolution simulated hosts. Alongside the need for more high resolution simulations, we also must account for distinct merger histories in shaping satellite-host system evolution. Building a statistical sample of high-resolution simulations with various merger histories will require ingenuity in both novel resolution enhancement techniques like particle splitting and re-simulation methods such as genetic modification. Developing robust methods of characterizing quenching mechanisms in both simulations and observations will also be essential to truly place Local Group satellites in a cosmological context by comparing to SAGA and an even larger sample of MW analogs that will be observable with the Nancy Grace Roman Space Telescope. The unique combination of next-generation simulations and the unparalleled observational detail achievable in the Local Group means that the future of galaxy formation is bright.

APPENDIX A

Chapter 2 Appendices

A.1. Resolution test

To examine the dependence of our satellite profiles on numerical resolution, we use the lower-resolution (LR) versions of the Latte simulation suite. We simulated each of the 7 Latte hosts at 8× lower mass resolution, with baryonic particle masses of $m_{bary} \sim 5.7 \times 10^4$ $M_\odot$ and $m_{dm} = 2.8 \times 10^5$ $M_\odot$. Furthermore, all gravitational force softenings are 2× larger.

In principle, we could compare satellites at fixed M_* between LR and high-resolution (HR) simulations. However, we choose to compare the survival of *subhalos* in the baryonic simulations that are resolved with the same number of DM particles, for three reasons. First, as studied extensively in Section 4.1.4 of Hopkins et al. (2018), the stellar masses of dwarf galaxies resolved with small numbers of star particles are sensitive to numerical convergence; the lowest-mass galaxies resolved in our LR simulations ($\sim$ 20 star particles) form systematically $\sim$ 2× higher M_* at fixed subhalo M_{peak} than in our HR simulations. Thus, comparing satellites at fixed M_* at our resolution limit mixes the numerical effects of star-formation efficiency and tidal disruption, but comparing satellites at fixed M_{peak} isolates the effects of tidal disruption, which is our goal here. Second, because dwarf galaxies at these masses are so DM-dominated, the survivability of a satellite is governed more directly by the number of DM particles in its subhalo than its number of star particles. Finally, most previous works on numerical disruption of satellites (e.g. van den Bosch & Ogiya, 2018) focused on DM-only simulations and how well resolved subhalos are, so using M_{peak} makes our tests more comparable to those previous works.

For the lowest-mass galaxies that we examine in this work, $M_* \sim 10^5$ $M_\odot$, we find that, across the HR simulations, they are hosted by subhalos with average $M_{peak} \sim 8 \times 10^8$ $M_\odot$ ($\sim 2 \times 10^4$ DM particles). Similarly, the HR satellites with $M_* \sim 10^6$ $M_\odot$ have subhalos of $M_{peak} \sim 2 \times 10^9$ $M_\odot$ ($\sim 7 \times 10^4$ DM particles) and HR satellites with $M_* \sim 10^7$ $M_\odot$ have subhalos of $M_{peak} \sim 10^{10}$ $M_\odot$ ($\sim 3 \times 10^5$ DM particles). With 8× larger particle mass in the LR simulations, a LR subhalo that is resolved as well as the subhalos of our lowest-mass satellites in the HR simulations has $M_{peak} \sim 6.4 \times 10^9$ $M_\odot$. We thus select subhalos in both the LR and HR simulations with $M_{peak} > 6.4 \times 10^9$ $M_\odot$ and compute their radial profiles out to 1000 kpc around each host, averaging over all 67 snapshots at $z = 0 - 0.1$ to improve statistics.

Figure A.1 shows a comparison of the LR radial profiles to the HR versions. The top panel compares the cumulative profile while the bottom panel compares the differential (discrete distance bins) profile, which more directly indicates where converge occurs. The solid lines and shaded regions show the median and 68 per cent and 95 per cent host-to-host scatter over the 7 Latte simulations. Within ∼30 kpc, the LR simulations show a deficit of ∼25 per cent (in the differential) compared with HR. Beyond 30 kpc, the median ratio stays mostly between 0.8 and 1, and the 68 per cent host-to-host scatter is always consistent with 1 for both cumulative and differential profiles. We thus conclude that the radial profiles are well converged (to ∼ 20 per cent) beyond ∼30 kpc, where almost all observed satellites ($M_* > 10^5$ $M_\odot$) of the MW and M31 are.

Moreover, the difference between LR and HR within 30 kpc is exaggerated by the fact that the LR simulations have more massive host galaxies. As with the dwarf galaxies themselves, and as studied in detail in Hopkins et al. (2018), the stellar masses of the host galaxies are sensitive to resolution as well, with LR hosts having on average 1.7 times higher stellar mass than their HR counterparts. Using the results from Figure 2.6 (left), a host galaxy with 1.7× higher stellar mass will have ∼15 per cent fewer satellites at d<50 kpc, even at

fixed resolution, which can account for most of the difference between LR and HR at small distances. Based on this, we plot the expected median values for the ratio of LR to HR profiles at 50 and 100 kpc as black points in the upper panel of Fig A.1. These profile ratio values are within 10 per cent of unity, showing better agreement between the LR and HR simulations.

Furthermore, given the complicating effects of different host galaxy masses in LR versus HR simulations, we also note similar results from the extensive numerical tests in Hopkins et al. (2018). Specifically, in Section 4.14 and Figure 14 they compared the (differential) number of subhalos versus distance in a DM-only simulation of the same m12i host at the same resolutions that we use here, using a broadly similar subhalo selection (instantaneous $M_{bound} > 10^8$ $M_\odot$). Thus, while that convergence test did not include the additional tidal force of the central galaxy or any other baryonic effects, its does provide a cleaner numerical test in the DM-only regime. They found convergence to better than 20 per cent down to $d \approx 50$ kpc, consistent with the results of Figure A.1.

We also can use the more rigorous criteria articulated in van den Bosch & Ogiya (2018) to test how well our lowest-mass subhalos are resolved. They stipulate that a subhalo on a circular orbit around a static, spherically symmetric host potential will suffer from numerical noise or disruption if the bound mass fraction falls below either of two limits that depend on mass and force resolution:

$$f_{\rm bound} < 0.32(N_{\rm acc}/1000)^{-0.8} \tag{A.1}$$

$$f_{\rm bound} < \frac{1.79}{f(c)} \left(\frac{\epsilon}{r_{\rm s,0}}\right) \left(\frac{r_{\rm h}}{r_{\rm s,0}}\right) \tag{A.2}$$

Where $N_{\rm acc}$ is the number of DM particles in the subhalo at accretion, c is the NFW concentration parameter of the subhalo at accretion, $f(c) = \ln(1+c) - \frac{c}{1+c}$, $r_{\rm s,0}$ is the NFW scale radius at accretion, and $r_{\rm h}$ is the instantaneous half-mass radius of the subhalo. We have verified that nearly all of the lowest mass subhalos considered in our HR simulations ($N_{\rm acc} \sim 2 \times 10^4$ DM particles) verify these criteria. Note however, that these criteria were generated from idealized DMO simulations that do not account for the disk potential present in our baryonic hosts, which more efficiently (and rapidly) disrupts subhalos that orbit close to the disk (see e.g. Garrison-Kimmel et al., 2017b).

A.2. Differential radial distribution

We also examined the differentially-binned radial profiles of satellites around the hosts in our simulations. Figure A.2 shows these profiles for the simulations and MW/M31 observations, considering all satellites with $M_* > 10^5$ $M_\odot$. The simulation scatter (blue regions) encompasses both the MW and M31 profiles at the 68 per cent level out to about 150 kpc. Past this distance, the MW shows a known lack of satellites between $150 - 200$ kpc, but is otherwise consistent with the simulations at the 68 per cent level. Potential incompleteness in the MW's satellite population is explored further in Section 2.5.8. M31 remains within the simulation scatter at the 95 per cent level from $150 - 300$ kpc. In general, the simulated differential distributions are a reasonable match to the Local Group.

A.3. Correlation with host galaxy and halo mass

In Figure A.3, we repeat the exercise of section 2.5.5, but this time plotting the number of satellites at small distances as a function of host halo mass and color coding my host disk mass. This illustrates that both host disk mass and halo mass simultaneously correlate with the number of satellites at small distances from the host. Therefore, the main driver of the negative trends with host mass remains uncertain in our analysis. However, other work that has systematically varied an analytical disk potential at fixed halo mass and found that the

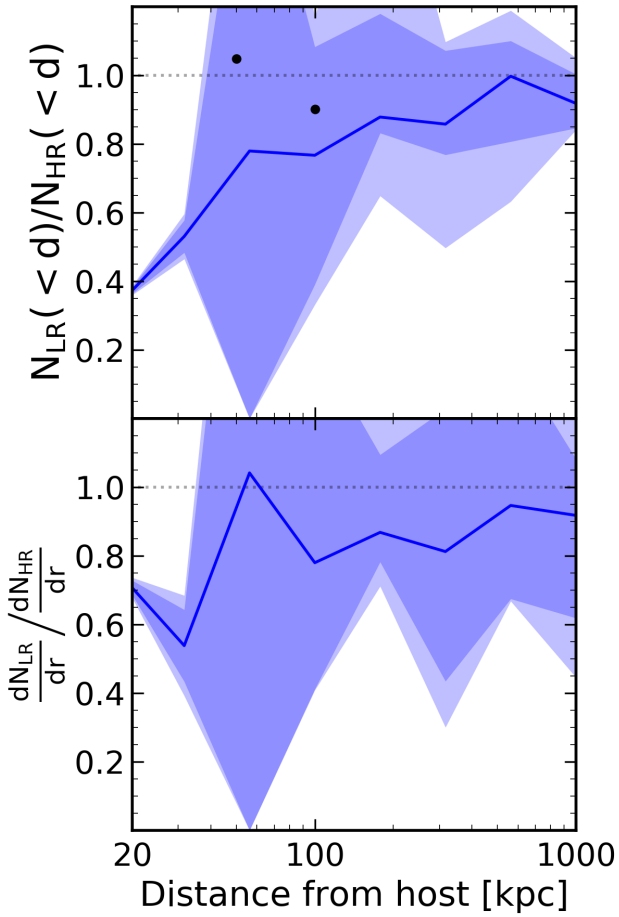

FIGURE A.1. Resolution test for subhalos with $M_{peak} > 6.4 \times 10^9$ $M_\odot$, which have the same number of DM particles in the low-resolution (LR) simulation as the lowest-mass subhalos that we analyze in the high-resolution (HR) simulations. *Top:* The cumulative radial profiles for subhalos in the LR simulations normalized to the HR profiles. The blue line shows the median and the shaded regions show the host-to-host scatter. The LR simulations have on average 80-100 per cent the number of subhalos as the HR simulations beyond about 30 kpc, indicating that we are resolving the satellites in our lowest mass bin from the main text. At small distances, the LR simulations have significantly fewer subhalos than the HR versions, but this is caused at least in part because of the more massive baryonic disks of the LR hosts. The black points represent an approximate model for removing this host mass effect at 50 and 100 kpc using the fits from Figure 2.6. *Bottom:* Same as top, but for differential radial binning instead of cumulative.

disk was the source of a reduction in DMO substructure close to the host when compared to a disk-less host halo (Kelley et al., 2019). Although our results are not necessarily definitive

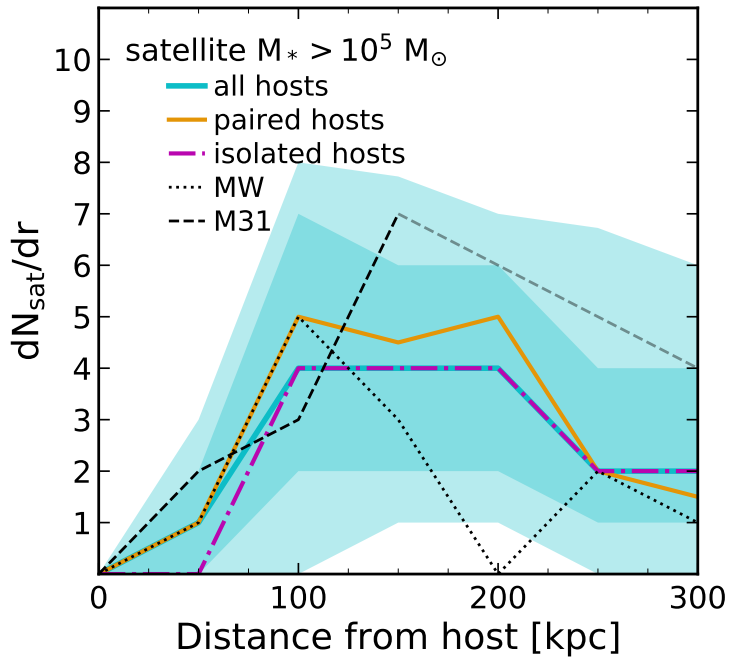

FIGURE A.2. Same as the top left panel of Figure 2.3 for satellites with $M_* > 10^5$ $M_\odot$, except the radial distribution uses differential bins instead of cumulative.

on their own, we conclude that it is not unreasonable for the trends we see in N_{sat} as a function of host mass to originate from enhanced tidal destruction of satellites due to the host's baryonic disk.

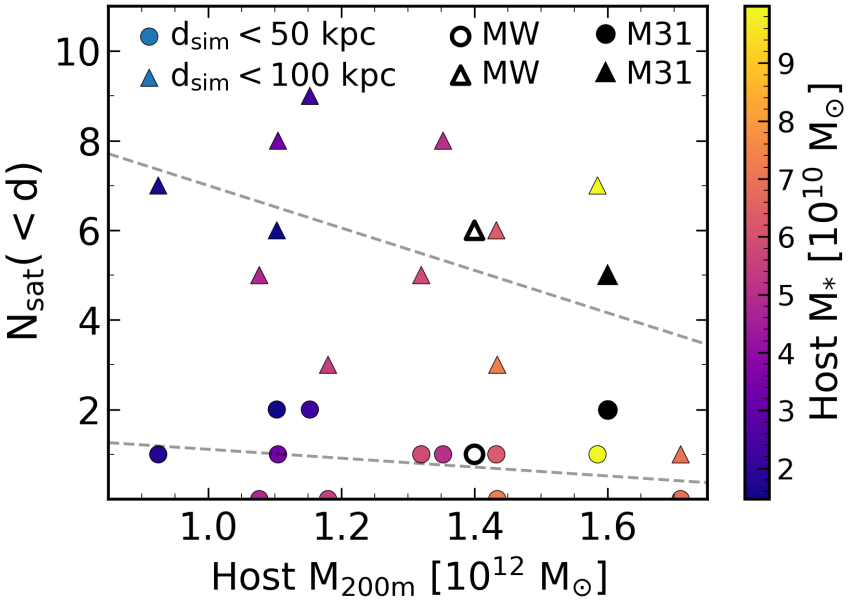

FIGURE A.3. Same as the right panel of Figure 2.6 for satellites with $M_* > 10^5$ $M_\odot$, but the points have been colored by the mass of the host galaxy's baryonic disk. Scatter has been left off of the points for clarity.

APPENDIX B

Chapter 3 Appendices

B.1. Alternative proper-motion measurements

Pawlowski & Kroupa (2020) compiled a 'best-available' sample of proper motions for their analysis of satellite planes by choosing measured proper motions of the 11 most massive classical dwarf MW satellites from the literature that have the smallest uncertainties. Compared to our sample, Pawlowski & Kroupa (2020) do not include Crater II, Antlia II, or Canes Venatici I in their sample, likely for consistency with past analyses and because the stellar masses of these satellites are close to the lower limit for classical dwarfs ($\sim 10^5 M_\odot$) and some of them show evidence for tidal disruption. Of the satellites that both of our samples have in common, all of the proper motions are approximately the same except for that of Leo II: Pawlowski & Kroupa (2020) uses a Leo II proper motion from Piatek et al. (2016) based on HST data, which is significantly different from the value we use in magnitude, direction, and uncertainty. Any other differences in proper motions between the two samples are not significant enough to alter our analysis.

Using the Piatek et al. (2016) proper motion and sampling from the given uncertainties, we measure a narrower range and lower median orbital pole dispersion (56 deg versus 60 deg) for MW satellites. Figure B.1 shows the main effect on our analysis: fewer simulation snapshots have MW-like orbital pole dispersions. Only 0.3 per cent of snapshots during $z = 0-0.2$ are at least as kinematically coherent as the MW satellite plane, whereas this value was previously 5 per cent using the Gaia proper motion for Leo II. Most importantly, the main conclusion of Section 3.6.4.1 still holds: MW-like planes are still more likely during the first pericentric passage of an LMC-like satellite, and using the best-available proper motion

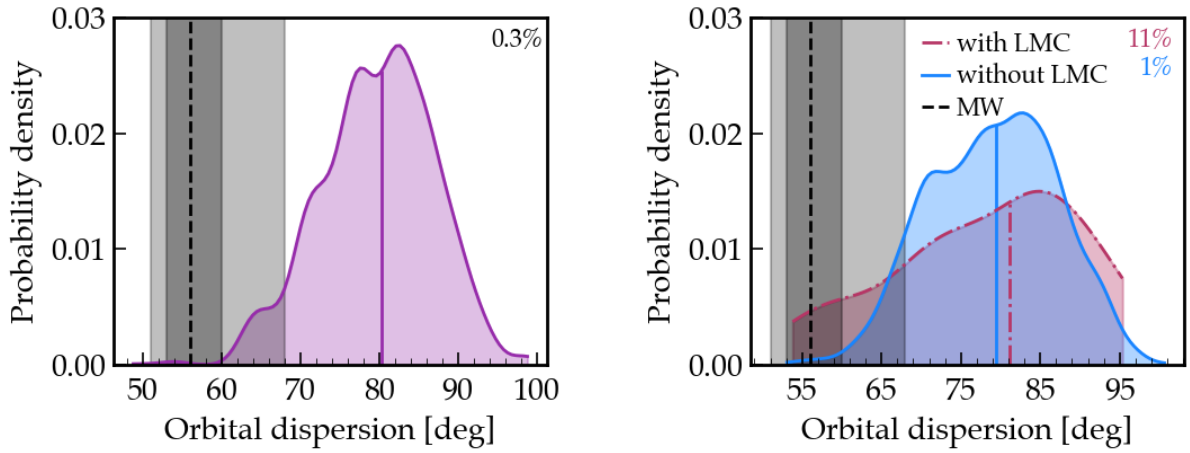

FIGURE B.1. Effects of an alternative proper motion sample for the observed MW satellites. Note that the redshift ranges for the left and right plots are not the same ($z = 0 - 0.2$ and $z \sim 0 - 0.7$, respectively). *Left:* Same as the right panel of Figure 3.2, but with the 'best-available' proper motion for Leo II from HST observations (Piatek et al., 2016). Using the best-available proper motion for Leo II shifts the median orbital pole dispersion for the MW from 60 deg to 56 deg, and also decreases the scatter to larger angles. This has the effect of reducing the fraction of snapshots with MW-like (at or below the MW's upper 68 per cent limit) planes from 5 per cent to 0.3 per cent, however, we note that almost 6 per cent of snapshots lie at or below the MW's upper 95 per cent limit. *Right:* Same as the right panel of Figure 3.7, but using the proper motions described above. While there are fewer snapshots that are MW-like overall, the enhancement in the fraction of snapshots that are MW-like during LMC-like pericenter passages relative to the general sample of snapshots is still evident.

sample actually enhances that result. Using the best-available proper motion sample, MW-like kinematic planes are 11 times more likely in the presence of our LMC analogues, versus $\sim$ 3 times more likely with our original proper motion sample. While the best-available proper motion sample makes MW-like kinematic planes more rare in our simulations, it does not affect our results on spatial thinness of planes and it does not qualitatively change any of our other results. A definitive observational proper motion for Leo II would enable us to perform an even more robust analysis of the MW's satellite plane in comparison to cosmological simulations.

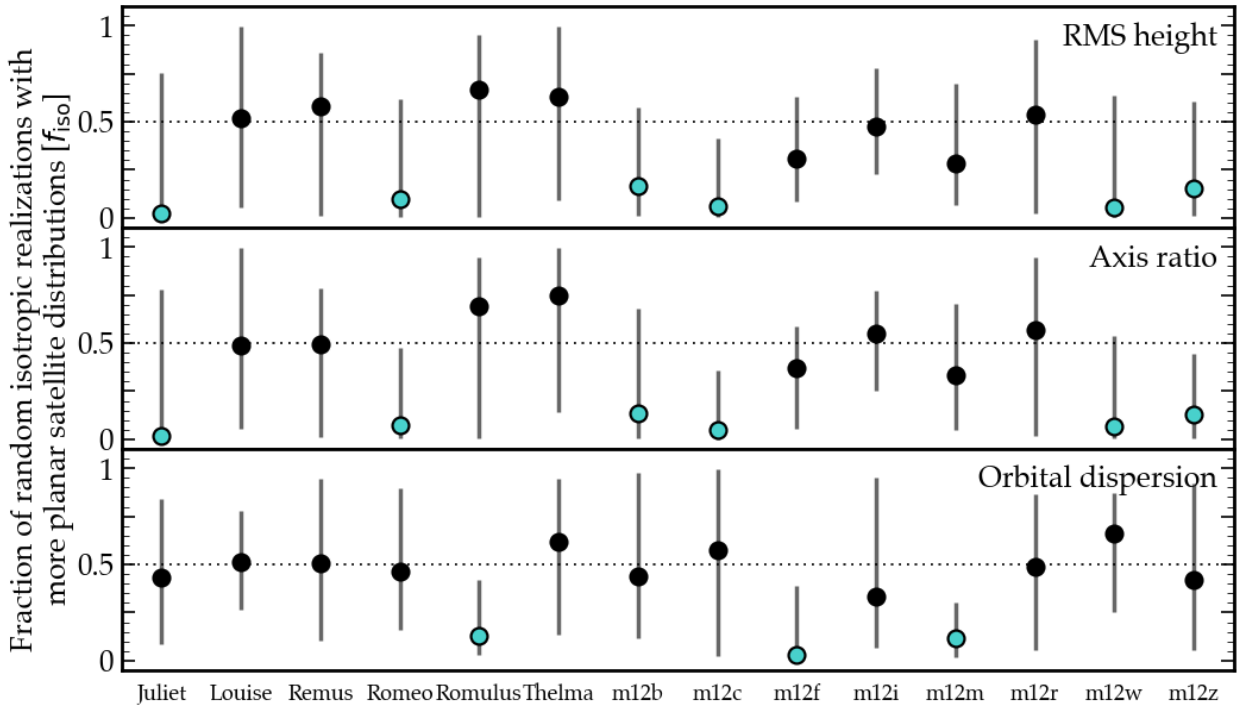

FIGURE B.2. The significance of simulated and observed plane metrics from Figure 3.2 relative to 10^4 statistically isotropic realizations of satellite positions (keeping radial distance fixed) and velocities (considering only their directions). For each simulated host we plot the median and 95 per cent scatter during $z = 0 - 0.2$ in the fraction ($f_{\rm iso}$) of isotropic realizations that are more planar than the true plane. We consider hosts with median $f_{\rm iso} \leq 0.25$ and lower 95 per cent limit $f_{\rm iso} \leq 0.05$ to have significant planes (blue). The MW's plane is highly significant relative to its statistically isotropic distribution, both spatially ($f_{\rm iso} = 0.003$) and kinematically ($f_{\rm iso} = 0.005$). About half of the simulated hosts (Juliet, Romeo, m12b, m12c, m12w, and m12z) have significant spatial planes, and only three (Romulus, m12f, and m12m) have significant kinematic planes during $z = 0 - 0.2$. None of the simulated hosts are significant in both a spatial and kinematic sense, and most hosts are consistent with a statistically isotropic distribution. While MW-like planes do occur in the simulations, they are not as significant as the MW's plane.

B.2. Comparison to statistically isotropic realizations

In Figure B.2, we provide a visual representation of how planar each host's satellite system is relative to 10^4 statistically isotropic random realizations of satellite positions and velocities. We describe this calculation in detail in Sections 3.5.3 and 3.6.2.1.

Bibliography

Ahmed S. H., Brooks A. M., Christensen C. R., 2017, MNRAS, 466, 3119

Akins H. B., Christensen C. R., Brooks A. M., Munshi F., Applebaum E., Engelhardt A., Chamberland L., 2021, ApJ, 909, 139

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, AJ, 156, 123

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

Bell E. F., de Jong R. S., 2001, ApJ, 550, 212

Belokurov V., et al., 2006, ApJ, 642, L137

Belokurov V., et al., 2007, ApJ, 654, 897

Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, MNRAS, 478, 611

Benítez-Llambay A., Navarro J. F., Abadi M. G., Gottlöber S., Yepes G., Hoffman Y., Steinmetz M., 2013, ApJL, 763, L41

Berezinsky V., Dokuchaev V., Eroshenko Y., 2006, PRD, 73, 063504

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

Bland-Hawthorn J., Gerhard O., 2016, Annual Review of Astronomy and Astrophysics, 54, 529

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, MNRAS, 422, 1203

Brainerd T. G., 2005, ApJL, 628, L101

Brainerd T. G., Samuels A., 2020, ApJL, 898, L15

Brooks A. M., Kuhlen M., Zolotov A., Hooper D., 2013, ApJ, 765, 22

Brown T. M., et al., 2012, ApJL, 753, L21

Buck T., Macciò A. V., Dutton A. A., 2015, ApJ, 809, 49

Buck T., Dutton A. A., Macciò A. V., 2016, MNRAS, 460, 4348

Buck T., Macciò A. V., Dutton A. A., Obreja A., Frings J., 2019, MNRAS, 483, 1314

Bullock J. S., Boylan-Kolchin M., 2017, AR&AA, 55, 343

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

Carlberg R. G., 1994, ApJ, 433, 468

Cautun M., Frenk C. S., 2017, MNRAS, 468, L41

Cautun M., Bose S., Frenk C. S., Guo Q., Han J., Hellwing W. A., Sawala T., Wang W., 2015, MNRAS, 452, 3838

Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C. A., Quataert E., 2015, MNRAS, 454, 2981

Chan T. K., Keres D., Hopkins P. F., Quataert E., Su K. Y., Hayward C. C., Faucher-Giguere C. A., 2018, arXiv e-prints, p. arXiv:1812.10496

Collins M. L. M., et al., 2013, ApJ, 768, 172

Collins M. L. M., et al., 2015, ApJL, 799, L13

Conn A. R., et al., 2012, ApJ, 758, 11

Conn A. R., et al., 2013, ApJ, 766, 120

D'Onghia E., Lake G., 2008, ApJL, 686, L61

D'Onghia E., Springel V., Hernquist L., Keres D., 2010, ApJ, 709, 1138

Deason A. J., Wetzel A. R., Garrison-Kimmel S., Belokurov V., 2015, MNRAS, 453, 3568

Dicke R. H., Peebles P. J. E., Roll P. G., Wilkinson D. T., 1965, ApJ, 142, 414

Diemand J., Kuhlen M., Madau P., 2007, ApJ, 667, 859

Dutton A. A., Macciò A. V., Frings J., Wang L., Stinson G. S., Penzo C., Kang X., 2016, MNRAS, 457, L74

Einasto J., Saar E., Kaasik A., Chernin A. D., 1974, Nature, 252, 111

El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, ApJ, 820, 131

Errani R., Peñarrubia J., Laporte C. F. P., Gómez F. A., 2017, MNRAS, 465, L59

Escala I., et al., 2018, MNRAS, 474, 2194

Fattahi A., Navarro J. F., Frenk C. S., Oman K. A., Sawala T., Schaller M., 2018, MNRAS, 476, 3816

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416

Fernando N., Arias V., Guglielmo M., Lewis G. F., Ibata R. A., Power C., 2017, MNRAS, 465, 641

Fernando N., Arias V., Lewis G. F., Ibata R. A., Power C., 2018, MNRAS, 473, 2212

Fillingham S. P., Cooper M. C., Pace A. B., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., Wheeler C., 2016, MNRAS, 463, 1916

Fitts A., et al., 2017, MNRAS, 471, 3547

Flores Velázquez J. A., et al., 2021, MNRAS, 501, 4812

Font A. S., et al., 2011, MNRAS, 417, 1260

Freedman W. L., et al., 2001, ApJ, 553, 47

Fritz T. K., Battaglia G., Pawlowski M. S., Kallivayalil N., van der Marel R., Sohn S. T., Brook C., Besla G., 2018, A&A, 619, A103

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

Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, ApJ, 757, 85

Geha M., et al., 2017, ApJ, 847, 4

Grcevich J., Putman M. E., 2009, ApJ, 696, 385

Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hargis J. R., Willman B., Peter A. H. G., 2014, ApJ, 795, L13

Harris C. R., et al., 2020, Nature, 585, 357

Hayashi E., Navarro J. F., Taylor J. E., Stadel J., Quinn T., 2003, ApJ, 584, 541

Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85

Hopkins P. F., 2015, MNRAS, 450, 53

Hopkins P. F., 2016, MNRAS, 455, 89

Hopkins P. F., et al., 2018, MNRAS, 480, 800

Hou A., Parker L. C., Harris W. E., 2014, MNRAS, 442, 406

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., et al., 2013, Nature, 493, 62

Ibata N. G., Ibata R. A., Famaey B., Lewis G. F., 2014a, Nature, 511, 563

Ibata R. A., Ibata N. G., Lewis G. F., Martin N. F., Conn A., Elahi P., Arias V., Fernando N., 2014b, ApJL, 784, L6

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., 2021, arXiv e-prints, p. arXiv:2106.03861

Jethwa P., Erkal D., Belokurov V., 2016, MNRAS, 461, 2212

Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org/

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

Kang X., Mao S., Gao L., Jing Y. P., 2005, A&A, 437, 383

Karunakaran A., et al., 2021, arXiv e-prints, p. arXiv:2105.09321

Kelley T., Bullock J. S., Garrison-Kimmel S., Boylan-Kolchin M., Pawlowski M. S., Graus A. S., 2019, MNRAS, 487, 4409

Kim S. Y., Peter A. H. G., Hargis J. R., 2018, PRL, 121, 211302

Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999a, ApJ, 516, 530

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999b, ApJ, 522, 82

Koposov S., et al., 2007, ApJ, 669, 337

Kravtsov A. V., Borgani S., 2012, AR&AA, 50, 353

Kroupa P., 2001, MNRAS, 322, 231

Kroupa P., Theis C., Boily C. M., 2005, A&A, 431, 517

Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36

Lam S. K., Pitrou A., Seibert S., 2015, in Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM '15. ACM, New York, NY, USA, pp 7:1–7:6, doi:10.1145/2833157.2833162, http://doi.acm.org/10.1145/2833157.2833162

Leitherer C., et al., 1999, The Astrophysical Journal Supplement Series, 123, 3

Li Y.-S., Helmi A., 2008, MNRAS, 385, 1365

Libeskind N. I., Cole S., Frenk C. S., Okamoto T., Jenkins A., 2007, MNRAS, 374, 16

Libeskind N. I., Frenk C. S., Cole S., Jenkins A., Helly J. C., 2009, MNRAS, 399, 550

Libeskind N. I., Knebe A., Hoffman Y., Gottlöber S., Yepes G., Steinmetz M., 2011, MNRAS, 411, 1525

Libeskind N. I., et al., 2020, MNRAS, 498, 2968

Lynden-Bell D., 1976, MNRAS, 174, 695

Lynden-Bell D., Lynden-Bell R. M., 1995, MNRAS, 275, 429

Macciò A. V., Kang X., Fontanot F., Somerville R. S., Koposov S., Monaco P., 2010, MNRAS, 402, 1995

Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, ApJ, 907, 85

Martin N. F., de Jong J. T. A., Rix H.-W., 2008, ApJ, 684, 1075

Martin N. F., et al., 2013a, ApJ, 772, 15

Martin N. F., Ibata R. A., McConnachie A. W., Mackey A. D., Ferguson A. M. N., Irwin M. J., Lewis G. F., Fardal M. A., 2013b, ApJ, 776, 80

Martin N. F., et al., 2016, ApJ, 833, 167

Mashchenko S., Wadsley J., Couchman H. M. P., 2008, Science, 319, 174

Mateo M. L., 1998, Annual Review of Astronomy and Astrophysics, 36, 435

McConnachie A. W., 2012, AJ, 144, 4

McConnachie A. W., et al., 2009, Nature, 461, 66

McGee S. L., Balogh M. L., Bower R. G., Font A. S., McCarthy I. G., 2009, MNRAS, 400, 937

Metz M., Kroupa P., Jerjen H., 2007, MNRAS, 374, 1125

Metz M., Kroupa P., Libeskind N. I., 2008, ApJ, 680, 287

Metz M., Kroupa P., Jerjen H., 2009, MNRAS, 394, 2223

Moore B., Katz N., Lake G., 1996, ApJ, 457, 455

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJL, 524, L19

Müller O., Pawlowski M. S., Jerjen H., Lelli F., 2018, Science, 359, 534

Nadler E. O., Mao Y.-Y., Wechsler R. H., Garrison-Kimmel S., Wetzel A., 2018, ApJ, 859, 129

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

Nelson D., et al., 2019, Computational Astrophysics and Cosmology, 6, 2

Neuzil M. K., Mansfield P., Kravtsov A. V., 2020, MNRAS, 494, 2600

Newton O., Cautun M., Jenkins A., Frenk C. S., Helly J. C., 2018, MNRAS, 479, 2853

Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092

Ostriker J. P., Tremaine S. D., 1975, ApJL, 202, L113

Pawlowski M. S., 2018, Modern Physics Letters A, 33, 1830004

Pawlowski M. S., Kroupa P., 2020, MNRAS, 491, 3042

Pawlowski M. S., McGaugh S. S., 2014, ApJL, 789, L24

Pawlowski M. S., Pflamm-Altenburg J., Kroupa P., 2012a, MNRAS, 423, 1109

Pawlowski M. S., Kroupa P., Angus G., de Boer K. S., Famaey B., Hensler G., 2012b, MNRAS, 424, 80

Pawlowski M. S., Kroupa P., Jerjen H., 2013, MNRAS, 435, 1928

Pawlowski M. S., et al., 2014, MNRAS, 442, 2362

Pawlowski M. S., Famaey B., Merritt D., Kroupa P., 2015, ApJ, 815, 19

Pawlowski M. S., et al., 2017, Astronomische Nachrichten, 338, 854

Pawlowski M. S., Bullock J. S., Kelley T., Famaey B., 2019, ApJ, 875, 105

Peñarrubia J., Benson A. J., Walker M. G., Gilmore G., McConnachie A. W., Mayer L., 2010, MNRAS, 406, 1290

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Pérez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21

Perlmutter S., et al., 1999, ApJ, 517, 565

Phillips J. I., Cooper M. C., Bullock J. S., Boylan-Kolchin M., 2015, MNRAS, 453, 3839

Piatek S., Pryor C., Olszewski E. W., 2016, AJ, 152, 166

Pillepich A., et al., 2018, MNRAS, 473, 4077

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06209

Pontzen A., Governato F., 2012, MNRAS, 421, 3464

Press W. H., Schechter P., 1974, ApJ, 187, 425

Putman M. E., Zheng Y., Price-Whelan A. M., Grcevich J., Johnson A. C., Tollerud E., Peek J. E. G., 2021, arXiv e-prints, p. arXiv:2101.07809

Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006a, MNRAS, 366, 429

Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006b, MNRAS, 367, 387

Riess A. G., et al., 1998, AJ, 116, 1009

Riley A. H., Strigari L. E., 2020, MNRAS, 494, 983

Rodriguez Wimberly M. K., Cooper M. C., Fillingham S. P., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., 2019, MNRAS, 483, 4031

Rubin V. C., Ford Jr. W. K., 1970, ApJ, 159, 379

Saha A., et al., 2010, AJ, 140, 1719

Sales L. V., Navarro J. F., Kallivayalil N., Frenk C. S., 2017, MNRAS, 465, 1879

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

Sanderson R. E., et al., 2018, arXiv e-prints, p. arXiv:1806.10564

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

Santos-Santos I., et al., 2020, ApJ, 897, 71

Sawala T., et al., 2016, MNRAS, 457, 1931

Sawala T., Pihajoki P., Johansson P. H., Frenk C. S., Navarro J. F., Oman K. A., White S. D. M., 2017, MNRAS, 467, 4383

Searle L., Zinn R., 1978, ApJ, 225, 357

Shao S., Cautun M., Frenk C. S., Grand R. J. J., Gómez F. A., Marinacci F., Simpson C. M., 2018, MNRAS, 476, 1796

Shao S., Cautun M., Frenk C. S., 2019, MNRAS, 488, 1166

Shaya E. J., Tully R. B., 2013, MNRAS, 436, 2096

Sick J., Courteau S., Cuilland re J.-C., Dalcanton J., de Jong R., McDonald M., Simard D., Tully R. B., 2015, in Cappellari M., Courteau S., eds, IAU Symposium Vol. 311, Galaxy Masses as Constraints of Formation Models. pp 82–85 (arXiv:1410.0017), doi:10.1017/S1743921315003440

Simpson C. M., Grand R. J. J., Gómez F. A., Marinacci F., Pakmor R., Springel V., Campbell D. J. R., Frenk C. S., 2018, MNRAS, 478, 548

Slater C. T., Bell E. F., 2014, ApJ, 792, 141

Smercina A., Bell E. F., Samuel J., D'Souza R., 2021, arXiv e-prints, p. arXiv:2107.04591

Sohn S. T., Patel E., Fardal M. A., Besla G., van der Marel R. P., Geha M., Guhathakurta P., 2020, ApJ, 901, 43

Spekkens K., Urbancic N., Mason B. S., Willman B., Aguirre J. E., 2014, ApJ, 795, L5

Spergel D. N., et al., 2003, ApJS, 148, 175

Springel V., 2005, MNRAS, 364, 1105

Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, MNRAS, 471, 144

Taylor J. E., Babul A., 2001, ApJ, 559, 716

Tollerud E. J., Bullock J. S., Strigari L. E., Willman B., 2008, ApJ, 688, 277

Tollerud E. J., et al., 2012, ApJ, 752, 45

Tollerud E. J., Boylan-Kolchin M., Bullock J. S., 2014, MNRAS, 440, 3511

Torrealba G., Koposov S. E., Belokurov V., Irwin M., 2016, MNRAS, 459, 2370

Torrealba G., et al., 2018, arXiv e-prints, p. arXiv:1811.04082

Walsh S. M., Willman B., Jerjen H., 2009, AJ, 137, 450

Weisz D. R., Dolphin A. E., Skillman E. D., Holtzman J., Gilbert K. M., Dalcanton J. J., Williams B. F., 2014, ApJ, 789, 148

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., 2013, MNRAS, 432, 336

Wetzel A. R., Deason A. J., Garrison-Kimmel S., 2015a, ApJ, 807, 49

Wetzel A. R., Tollerud E. J., Weisz D. R., 2015b, ApJ, 808, L27

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23

Woo J., Courteau S., Dekel A., 2008, MNRAS, 390, 1453

Yniguez B., Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., 2014, MNRAS, 439, 73

Zabludoff A. I., Mulchaey J. S., 1998, ApJ, 498, L5

Zentner A. R., Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2005, ApJ, 629, 219

Zhu Q., Marinacci F., Maji M., Li Y., Springel V., Hernquist L., 2016, MNRAS, 458, 1559

Zwicky F., 1937, ApJ, 86, 217

van Kampen E., 1995, MNRAS, 273, 295

van Kampen E., 2000, arXiv e-prints, pp astro–ph/0002027

van den Bosch F. C., Ogiya G., 2018, MNRAS, 475, 4066

van der Marel R. P., Kallivayalil N., 2014, ApJ, 781, 121

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