Fall 10-1-2021

The tidal evolution of dark matter substructure and the significance of halo-to-halo assembly history variance

Sheridan Beckwith Green
Yale University Graduate School of Arts and Sciences, sheridan.b.green@gmail.com

Abstract

The Tidal Evolution of Dark Matter Substructure and the Significance of Halo-to-Halo Assembly History Variance

Sheridan B. Green

2021

Accurately predicting the abundance and structural evolution of dark matter subhaloes is crucial for understanding galaxy formation, large-scale structure, and constraining the nature of dark matter. Due to the nonlinear nature of subhalo evolution, cosmological N-body simulations remain its primary method of investigation. Subhaloes in such simulations have recently been shown to still be heavily impacted by artificial disruption, diminishing the information content (at small scales) of the simulations and all derivative semi-analytical models calibrated against them.

A model of the evolved subhalo density structure Our recent release of the *DASH* library of high-resolution, idealized N-body simulations of the tidal evolution of subhaloes (unhindered by numerical over-merging due to discreteness noise or force softening) enables a more accurate calibration of semi-analytical treatments of dark matter substructure evolution. We use *DASH* to calibrate a highly accurate, simply parametrized empirical model of the evolved subhalo density profile (ESHDP), which captures the impact of tidal heating and stripping. By testing previous findings that the structural evolution of a tidally truncated subhalo depends solely on the fraction of mass stripped, independent of the details of the stripping, we identify an additional dependence on the initial subhalo concentration. We provide significantly improved fitting functions for the subhalo density profiles and structural parameters ($V_{\rm max}$ and $r_{\rm max}$) that are unimpeded by numerical systematics and applicable to a wide range of parameter space.

A model of the build-up and evolution of dark matter substructure By combining our ESHDP model with a physically motivated prescription for the subhalo mass stripping rate, we introduce a state-of-the-art model of the mass evolution of individual subhaloes. This model has been calibrated to reproduce the mass trajectories of subhaloes in the *DASH* simulations. We incorporate this treatment of the subhalo internal structure and mass evolution into the recently released `SatGen` semi-analytical model. `SatGen` combines (i) analytical halo merger trees, (ii) a recipe for initial subhalo orbits at infall, (iii) an orbit integrator (which captures dynamical friction), and (iv) our *DASH*-calibrated tidal evolution model in order to ultimately capture the build-up and evolution of populations of dark matter substructure. We also develop a model of artificial disruption that reproduces the statistical properties of disruption in the *Bolshoi* simulation. Using

the *DASH*-calibrated `SatGen` framework, we generate independent predictions for key quantities in small-scale cosmology, including the evolved subhalo mass function, subhalo radial abundance, and the substructure mass fraction and study how these quantities are impacted by artificial disruption and mass resolution limits. We find that artificial disruption affects these quantities at the $10-20\%$ level, ameliorating previous concerns that it may suppress the SHMF by as much as a factor of two. We demonstrate that semi-analytical substructure modeling must include orbit integration in order to properly account for splashback haloes, which make up roughly half of the subhalo population. We show that the resolution limit of N-body simulations, rather than artificial disruption, is the primary cause of the radial bias in subhalo number density found in dark matter-only simulations. Hence, we conclude that the mass resolution remains the primary limitation of using such simulations to study subhaloes.

The impact of a galactic disc on the subhalo population Numerical simulations have shown that the formation of a central disc can drastically reduce the abundance of substructure compared to a dark matter-only simulation, which has been attributed to enhanced destruction of substructure due to disc shocking. We examine the impact of discs on substructure using `SatGen`. Using a sample of 10,000 merger trees of Milky-Way like haloes, we study the demographics of subhaloes that are evolved under a range of composite halo–disc potentials with unprecedented statistical power. We find that the overall subhalo abundance is relatively insensitive to properties of the disc aside from its total mass. For a disc that contains 5% of $M_{\rm vir}$, the mean subhalo abundance within $r_{\rm vir}$ is suppressed by $\lesssim$10% relative to the no-disc case, a difference that is dwarfed by halo-to-halo variance. For the same disc mass, the abundance of subhaloes within 50 kpc is reduced by $\sim$30%. We argue that the disc mainly drives excess mass loss for subhaloes with small pericentric radii and that the impact of disc shocking is negligible.

The three subhalo-focused studies described above constitute the primary thrust of this dissertation. However, the analytical Monte Carlo merger tree method, which is a key component of `SatGen`, has additional utility beyond the realm of subhalo studies. Indeed, an overarching theme of this program is that variation in assembly histories propagates to substantial halo-to-halo variance in *many* quantities of astrophysical and cosmological interest. We expand on this motif in the following two studies.

The impact of assembly history variance on cluster scaling relations X-ray and microwave cluster scaling relations are immensely valuable for cosmological analysis. However, their power is limited by astrophysical systematics that bias mass estimates and introduce additional scatter. Turbulence injected into the intracluster medium via mass assembly contributes substantially to cluster non-thermal pressure support, a significant source of such uncertainties. We use an analytical model to compute the assembly-driven non-thermal pressure profiles of haloes based on Monte Carlo-generated accretion histories (leveraging

the same method that is used to generate merger trees in `SatGen`). We introduce a fitting function for the average non-thermal pressure fraction profile, which exhibits minimal dependence on redshift at fixed peak height. Using the model, we predict deviations from self-similarity and the intrinsic scatter in the Sunyaev–Zel'dovich effect observable-mass scaling relation ($Y_{\rm SZ} - M$) due solely to inter-cluster variation in mass accretion histories. We study the dependence of $Y_{\rm SZ} - M$ on aperture radius, cosmology, redshift, and mass limit. The model predicts $5 - 9\%$ scatter in $Y_{\rm SZ} - M$ at $z = 0$, increasing as the aperture used to compute $Y_{\rm SZ}$ increases from $R_{\rm 500c}$ to $5R_{\rm 500c}$. The predicted scatter lies slightly below that of studies based on non-radiative hydro-simulations, illustrating that assembly history variance is likely responsible for a substantial fraction of scatter in $Y_{\rm SZ} - M$. This should be regarded as a lower bound, which will likely increase with the use of an updated gas density model that incorporates a more realistic response to halo assembly. As redshift increases, $Y_{\rm SZ} - M$ deviates more from self-similarity and scatter increases. We show that the $Y_{\rm SZ} - M$ residuals correlate strongly with the recent halo mass accretion rate, potentially providing an opportunity to infer the latter.

Estimating cluster masses via machine learning We present a machine-learning approach for estimating galaxy cluster masses, trained using both *Chandra* and *eROSITA* mock X-ray observations of 2041 clusters from the *Magneticum* simulations. We train a random forest (RF) regressor, an ensemble learning method based on decision tree regression, to predict cluster masses using an input feature set. The feature set uses core-excised X-ray luminosity and a variety of morphological parameters, including surface brightness concentration, smoothness, asymmetry, power ratios, and ellipticity. The regressor is cross-validated and calibrated on a training sample of 1615 clusters (80% of sample), and then results are reported as applied to a test sample of 426 clusters (20% of sample). This procedure is performed for two different mock observation series in an effort to bracket the potential enhancement in mass predictions that can be made possible by including dynamical state information. The first series is computed from idealized *Chandra*-like mock cluster observations, with high spatial resolution, long exposure time (1 Ms), and the absence of background. The second series is computed from realistic-condition *eROSITA* mocks with lower spatial resolution, short exposures (2 ks), instrument effects, and background photons modeled. We report a 20% reduction in the mass estimation scatter when *either series* is used in our RF model compared to a standard regression model that only employs core-excised luminosity. The morphological parameters that hold the highest feature importance are smoothness, asymmetry, and surface brightness concentration. Hence these parameters, which encode the dynamical state of the cluster, can be used to make more accurate predictions of cluster masses in upcoming surveys, offering a crucial step forward for cosmological analyses.

The Tidal Evolution of Dark Matter Substructure and the Significance of Halo-to-Halo Assembly History Variance

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Sheridan B. Green

Dissertation Director: Prof. Frank C. van den Bosch

December 2021

For my brothers, my parents, and April

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my advisor, Frank van den Bosch. Prior to taking his course on galaxy formation and getting to know him, I was strongly considering leaving astrophysics and potentially graduate school altogether. He has since become a constant force of inspiration; indeed, his boundless enthusiasm for science has greatly influenced the success of my research career to date. He has been a wonderful mentor — nearly always available but sufficiently hands-off to allow me to develop independence. I thank him immeasurably for continuing to support me after I made the decision to pursue a career in finance after the PhD.

I am also grateful to Daisuke Nagai for his mentorship and support of my interest in data science and statistics. Thanks to him, I had the chance to develop a machine learning background and work with Michelle Ntampaka, perhaps the most encouraging scientist I have ever met. Daisuke also facilitated my initial work with Jessi Cisewski-Kehe, who introduced me to the fascinating field of topological data analysis and also helped me along as I began my own journey as a mentor.

I would like to thank Adrienne Erickcek for being a wonderful undergraduate thesis advisor. During our work together, I realized just how easy it is to be led astray by numerical systematics when studying N-body simulation results, an insight that ultimately led me to Frank's research. I am also immensely grateful to Donald Jacobs for facilitating my first collaborative undergraduate research experience, which confirmed my interest in pursuing a scientific career in the first place.

I am thankful to John Murray and Josh Burt for giving me a chance to pretend I was a neuroscientist while I was searching for new academic paths. I would also like to thank my collaborators — Fangzhou Jiang, Go Ogiya, Han Aung, Tim Miller, and Xin Xu — as well my group members — Uddipan Banik, Dhruba Dutta Chowdury, Nir Mandelker, Kaustav Mitra, and Michael Tremmel.

Lastly, I extend an uncountable infinity of thanks to my friends, family, and April. To my friends and brothers — the time spent together, often on Discord, over the past few years has been nothing short of therapeutic. To my parents — it would be impossible to overstate the amount of love and encouragement you two have given me since the beginning. And to April — thanks for going on this journey with me.

Chapter 1

Introduction

1.1 Cosmological structure formation

The cosmological footprint of dark matter was first identified nearly 90 years ago during Fritz Zwicky's study of the virialization of the Coma cluster (Zwicky, 1933), where he noted that the cluster had an extremely insufficient stellar mass relative to the amount necessary to keep the system gravitationally bound. Like many revolutionary paradigm shifts before, it took the astronomy community quite some time to accept the existence of dark matter. Further work by Ostriker & Peebles (1973) and others in the 1970s demonstrated that, indeed, even the behavior of satellite galaxies around ordinary galaxies such as the Milky Way demands the existence of a massive dark halo.

Several decades of theoretical work coupled with exquisite measurements of temperature fluctuations in the cosmic microwave background (CMB) by the *COBE*, *WMAP*, and *Planck* satellites (e.g., Mather et al., 1990; Spergel et al., 2003; Planck Collaboration et al., 2014a) have since cemented the presence of dark matter in our Universe. The mass ratio between baryonic matter (i.e., all matter coupled to photons) and dark matter is roughly 1:5, making dark matter a substantial fraction of the cosmic energy density. Many models have been proposed to explain the nature of dark matter, ranging from massive compact halo objects (MACHOs; e.g., Alcock et al., 2000) that simply include baryonic objects that are too faint to see (such as brown dwarfs and black holes) to more exotic solutions such as modified Newtonian dynamics (MOND; e.g., Milgrom, 1983), which varies the relationship between force and gravitational acceleration as a function of the local density. However, thanks to contemporaneous progress in theoretical and experimental particle physics, the majority of the research community has since settled down on the hypothesis that dark matter

is composed of an elementary particle, several favorites of which currently include sterile neutrinos (for a review, see Boyarsky et al., 2019), weakly interacting massive particles (WIMPs; see Jungman et al., 1996), and ultra-light axion-like bosons (Hui et al., 2017), among others. These particles are separated into categories (cold, warm, and hot) based on their free-streaming velocities (i.e., the thermal velocity at freeze-out for thermally produced particles).

Several outstanding issues still remain in particle physics and the proposed particles that solve these problems have varying potential to also serve as a dark matter candidate. For example, the existence of a sterile neutrino of sufficient mass to simultaneously serve as dark matter would naturally explain the neutrino mass hierarchy. The presence of axion dark matter would simultaneously solve the strong CP problem. A WIMP, which is usually expected to be the lightest supersymmetric partner (typically a neutralino) or the lightest Kaluza-Klein particle in universal extra dimension theories, would help solve the hierarchy problem and potentially identify the best beyond-the-Standard Model (BSM) theory to continue pursuing. Recently, there has been limited progress in experimental particle physics towards successfully detecting any such predicted dark matter candidate. Fortunately, however, astronomical observations can be used to constrain the properties of dark matter structure, which in turn can also be used to constrain its underlying nature.

In the standard cosmological model, the primordial density field, $\rho(\mathbf{x})$, is seeded with (potentially Gaussian) initial density perturbations, which are described by the overdensity field, $\delta(\mathbf{x}) = (\rho(\mathbf{x}) - \bar{\rho})/\bar{\rho}$, where $\bar{\rho}$ is the background energy density, which is the same as the critical density in a flat cosmology (e.g., Dodelson, 2003). The initial power spectrum of these perturbations is nearly scale-invariant, i.e., $P_{\rm i}(k) = \langle|\delta(\mathbf{k})|^2\rangle \propto k^n$, $n \approx 1$, k is the wavenumber, and $\delta(\mathbf{k})$ is the Fourier transform of $\delta(\mathbf{x})$. The initial perturbations are very small in amplitude and, thus, their early evolution can be modeled independently using a linear approximation. Once a particular mode enters the cosmic horizon, it undergoes dynamical evolution according to, schematically,

$$\frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} + [\text{Pressure} - \text{Gravity}]\delta = 0. \tag{1.1}$$

Pressure and gravity are at constant odds with one another. When gravity dominates, overdensities experience growth; when pressure wins, δ undergoes oscillations. The pressure term is significant for baryons but is negligible for cold dark matter (CDM). The dominant component of the cosmic energy density controls the *rate* at which overdensities grow, with $\delta \propto D(t)$. In the early radiation-dominated era, $D(t) \propto \ln[a(t)]$, where $a(t)$ is the scale factor. During the period of matter domination, modes grow much more rapidly, with $D(t) \propto a(t)$. Because of this, the transition from logarithmic to linear growth at the epoch of matter–radiation equality impresses a characteristic feature into the power spectrum at $k \approx k_{\rm eq}$, the horizon scale

at equality. Furthermore, Silk damping (for baryons; Silk, 1968) and free-streaming damping (for collisionless dark matter; negligible for CDM) suppress power on small scales. These scale-dependent effects are captured in the "transfer function", $T(k)$. Thus, the linearly evolved power spectrum is related to the initial power spectrum according to

$$P(k,t) = P_{\rm i}(k)T^2(k)D^2(t). \tag{1.2}$$

The shape of the primordial power spectrum, $P_{\rm i}(k)$, is predicted by a model of the early Universe. The most widely accepted family of such models is based on cosmic inflation (Guth & Pi, 1982), during which the scale factor experiences exponential growth that causes quantum fluctuations to become permanently impressed into the highly stretched density field. Linear theory successfully captures the subsequent evolution of the perturbations after inflation. However, the *normalization* of $P_{\rm i}(k)$, and thus $P(k,t)$, is not predicted by the model and must be fit to observations. Typically, the observable used in order to properly set this normalization is a proxy for the variance of the density field within randomly placed spheres of radii R, which we write as

$$\sigma^2(R) = \frac{1}{2\pi^2}\int P(k)\hat{W}_R^2(k)k^2{\rm d}k \tag{1.3}$$

(e.g., Mo et al., 2010). Here, $\hat{W}_R(k)$ is the Fourier transform of the spherical top-hat window function (i.e., a step function with constant, nonzero value only for $r \leq R$). By convention, R is taken to be $8h^{-1}$ Mpc such that $\sigma_8 \equiv \sigma(R = 8h^{-1}\ {\rm Mpc})$ sets the normalization. Observations of the spatial distribution of galaxies and of temperature fluctuations in the CMB provide two different means to measure σ_8, which is of order unity.

Once the perturbations grow to be sufficiently large, linear theory begins to break down. At this point, higher-order perturbation theories (e.g., Peebles, 1980; Bernardeau, 1994) can be of some limited use. However, the advent of high-performance computing systems over the past several decades has enabled the use of computationally expensive numerical approaches, primarily the N-body simulation. The linear power spectrum, in conjunction with the Zel'dovich approximation (Zel'Dovich, 1970), is used to generate a random set of initial positions and velocities for the dark matter particles in the simulation (at a high redshift, $z_{\rm i}$). These particles then undergo gravitational evolution until $z = 0$ such that their final spatial distribution can be compared to observational proxies for the dark matter distribution. Cosmological N-body simulations have proven to be incredibly powerful for placing constraints on dark matter models and, more broadly, models of cosmology. In particular, comparisons between the small-scale clustering of dark matter (from simulations) and observed clustering of galaxies has placed strong constraints on the thermal velocity of the particle (see e.g., Primack & Gross, 2001; Lovell et al., 2012), ruling out most "hot" and some "warm" dark matter models and placing preference towards CDM. Discrepancies between simulations and galaxy surveys on large scales (Efstathiou et al., 1990) signaled the need for an additional component in

the cosmological model. Indeed, after Type Ia supernovae observations demonstrated that the Universe is expanding at an accelerating rate (Riess et al., 1998), a cosmological constant (Λ, also known as "dark energy") was incorporated into the *standard* model of cosmology. In this Λ cold dark matter (ΛCDM) model, structure forms hierarchically as self-bound clumps of dark matter (*haloes*) merge together to form ever larger systems. We elaborate on the consequences of this assembly process below.[1]

1.2 Dark matter haloes

After a density perturbation grows sufficiently large in magnitude ($\delta \approx 1$), it enters the non-linear regime. We can explore the subsequent evolution, which eventually proceeds to gravitational collapse, using the idealized "spherical collapse model" (Gunn & Gott, 1972). This model follows the evolution of a spherical top-hat overdensity in an expanding Einstein–de Sitter (EdS) universe (i.e., matter-dominated and flat). By studying the equation of motion of this perturbation, it can be shown that the top-hat initially expands until it reaches a "turn-around radius" and then subsequently undergoes gravitational collapse (i.e., approaches a singularity with infinite overdensity). The most useful component of the spherical collapse model is its prediction of the *time* of collapse. This time can be plugged in to the linear theory prediction of perturbation growth (i.e., $\delta(t) \propto D(t) \propto a(t) \propto t^{2/3}$ in the matter-dominated era) in order to demonstrate that regions in the *linearly extrapolated* density field with $\delta \geq \delta_c = 1.686$ should have undergone non-linear gravitational collapse (see e.g., Mo et al., 2010). This approach can be used to estimate the time (or redshift) of collapse (t_c) as well, since a perturbation should begin to undergo collapse when $\delta(t_c) = \delta_c D(\text{today})/D(t_c)$. In reality, the linear theory-based estimate of the critical overdensity for collapse (δ_c) drastically underestimates the true overdensity of a non-linearly collapsed object. A simple application of the virial theorem to the result of the spherical collapse model can be used to show that dark matter overdensities collapse to form virialized objects with $\Delta_{\rm vir} \equiv \delta(t_c) \approx 178$ in an EdS cosmology. In a ΛCDM cosmology (i.e., flat with roughly 70% dark energy and 30% matter), the virial overdensity is around $\Delta_{\rm vir} \approx 100$ (Bryan & Norman, 1998). These quasi-equilibrium objects are referred to as *haloes*. The boundary of the halo is typically defined to be the radius, $r_{\rm vir}$, within which the mean density is equal to $\Delta_{\rm vir}$ times the critical density — the mass enclosed within this radius is referred to as the "virial mass" of the halo, $M_{\rm vir}$.

The formation of CDM haloes proceeds in a hierarchical fashion: small objects form first, with larger systems forming as the product of mergers and the "smooth accretion" of previously unbound material. In 1974, (Press & Schechter, 1974) presented a statistical model of halo abundance, which counts the regions

1. The studies contained within this dissertation build directly off of the PhD work of Jiang (2016). As such, we draw heavily upon the introductory material of Jiang (2016) and delve into advancements that have been made in the field since 2016 when relevant.

in an initially Gaussian overdensity field linearly extrapolated to today with $\delta \geq \delta_c$ (from the spherical collapse model). The abundance is determined as a function of halo mass by counting the collapsed regions of the *smoothed* density field (using a smoothing scale, R, that corresponds to the halo mass, M). The Press–Schechter mass function, which gives the comoving number density of haloes between mass M and $M+\mathrm{d}M$, is written as

$$n(M,t)\,\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M^2}\frac{\delta_c(t)}{\sigma(M)}\exp\left(-\frac{\delta_c^2(t)}{2\sigma^2(M)}\right)\left|\frac{\mathrm{d}\ln\sigma(M)}{\mathrm{d}\ln M}\right|\mathrm{d}M, \tag{1.4}$$

which we emphasize only depends on time via $\delta_c(t) \equiv 1.686/D(t)$ (normalized such that $D(t) = 1$ today) and only depends on halo mass via $\sigma(M)$. Note that $\sigma(M)$ is simply $\sigma(R)$, where R is the radius of the Lagrangian volume corresponding to M; i.e., $M = \frac{4}{3}\pi R^3\bar{\rho}$. By introducing the halo "peak height", $\nu \equiv \delta_c(t)/\sigma(M)$, one can show that the mass function is purely a function of ν, which absorbs all dependence on time, mass, cosmology, and power spectrum into one variable.

The Press–Schechter formalism was extended by Bond et al. (1991) and Lacey & Cole (1993) into what is now known as, conveniently, the extended Press–Schechter (EPS) theory. The EPS framework, based on excursion set formalism, provides the necessary firepower to compute the *conditional* mass function (CMF), $f(M_1|M_2)$, which expresses the fraction of mass in haloes of mass M_2 at t_2 that is contributed by progenitor haloes of mass $M_1 < M_2$ at an earlier time $t_1 < t_2$. The CMF serves as the basis of (semi-)analytical Monte Carlo algorithms that have been developed to generate random halo mass accretion histories (MAHs; e.g., Cole et al., 2000; Parkinson et al., 2008); these MAHs are used as a component in a variety of modeling frameworks (see Section 1.3.3). The method is referred to as *semi*-analytical because the overall merger rate is typically calibrated to reproduce the halo statistics of cosmological simulations. The statistics of these random MAH realizations play a central role in this dissertation. As we will show, the large halo-to-halo variance in MAHs (at fixed halo mass) propagates to large halo-to-halo variance in a wide range of statistics of astrophysical significance.

While much can be accomplished via semi-analytical modeling techniques (and indeed, such is the focus of this dissertation), a large fraction of that which has been learned about dark matter haloes over the past few decades is attributed to cosmological N-body simulations, such as the *Aquarius* (Springel et al., 2008) and *Bolshoi* (Klypin et al., 2011) simulations. For example, simulations have been used to show that CDM haloes have a nearly universal structure, with a spherically-averaged density profile of

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1+\frac{r}{r_s}\right)^2} \tag{1.5}$$

(the NFW profile, named after Navarro et al., 1997), although we note that other profiles exist in the liter-

ature (e.g., Einasto, 1965; Dekel et al., 2017). Here, r_s is the "scale radius", which characterizes the region where the log-slope of the density profile transitions from -1 to -3, and ρ_s is the associated scale density. The (r_s, ρ_s) can be mapped to (M_{vir}, c), where M_{vir} is the halo virial mass and $c \equiv r_{vir}/r_s$ is the *concentration*. Here, $\rho_s = \frac{M_{vir}}{4\pi r_s^3}[\ln(1+c) - c/(1+c)]^{-1}$. The concentration is closely connected to halo age, and thus halo mass, which results in a fairly tight concentration–mass relation (e.g., Prada et al., 2012; Dutton & Macciò, 2014) and concentration–formation time relation (e.g., Wechsler et al., 2002; Zhao et al., 2009). The reason for such a connection is simple: less massive haloes form earlier when the background density is higher. As a consequence, the central densities of smaller haloes must be larger, with more of the total mass concentrated near the halo center at present day. Hence, c tends to increase as M_{vir} decreases.

The sophistication of most semi-analytical models stops here — haloes are often assumed to be smooth, spherically symmetric objects with no angular momentum. For many applications, these are reasonable assumptions to make. However, cosmological simulations have demonstrated that haloes are typically *not* spherically symmetric but are instead triaxial (e.g., Jing & Suto, 2002). Indeed, observations appear to confirm that, while low-mass groups tend to be spherical, high-mass groups are often more prolate (e.g., Wang et al., 2008). The formation of massive haloes is a rather anistropic process; mass is preferentially fed to the halo along cosmic filaments (e.g., Libeskind et al., 2013). Haloes also tend to have non-negligible spin (i.e., angular momentum) that is also associated with its position in the cosmic web (e.g., Aragón-Calvo et al., 2007). Lastly, due to the hierarchical assembly process of CDM haloes, these systems are abundant with *substructure*, the central focus of this dissertation.

1.3 Halo substructure

1.3.1 Evolution processes

When two haloes merge, the remnant of the smaller body lives on as a *subhalo*, orbiting within its *host* as it is subjected to destructive forces. However, it was not until the late 1990s that cosmological simulations achieved sufficient resolution to consistently resolve these subhaloes (e.g., Ghigna et al., 1998; Tormen et al., 1998). Further increases in computational power have enabled a wealth of studies aimed at quantifying the substructure abundance in cosmological simulations (e.g., Springel et al., 2008; Giocoli et al., 2010; Klypin et al., 2011; van den Bosch, 2017). These studies have reported relatively consistent results: the subhalo mass function (SHMF), which describes the number of subhaloes (of mass m) within a host (of mass M) per logarithmic mass bin, tends to follow $\mathrm{d}N/\mathrm{d}\ln(m/M) \propto (m/M)^{-(0.7-1.0)}$. The SHMF is expected to follow such a power law down to the free-streaming mass, which is roughly an Earth mass for typical CDM models

and is much larger for WDM ($\sim 10^9\,M_\odot$; e.g., Lovell et al., 2014). The power law slope of the SHMF is a result of (i) the nearly scale-invariant power spectrum, which sets the shape of the "field" halo mass function, and (ii) the dynamical processes undergone by subhaloes post-merger. These processes include tidal stripping, tidal heating, and dynamical friction due to the host. We elaborate on these processes below. However, we note that two additional second-order processes, which we will not discuss further, include tidal heating via impulsive encounters with other subhaloes (referred to as *harassment*) and self-friction due to previously stripped subhalo material.

An orbiting subhalo experiences mass loss via *tidal stripping*, a process by which material that lies beyond the Hill surface of the subhalo is peeled off and transferred to the host. In reality, this zero-acceleration surface is not spherical (e.g., Tollet et al., 2017). However, models of the "tidal radius", which assumes a spherical Hill surface, are often quite accurate. The simplest example assumes two point masses, m (subhalo) and M (host halo), separated by a distance of R. The Roche limit tidal radius, r_t, of m is the distance from m where the tidal force from M overtakes the self-gravity of m, which is $r_\mathrm{t} = R(\frac{m}{2M})^{1/3}$. After generalizing to non-circular orbits and extended mass profiles for the host and subhalo, King (1962) provides an approximation of the tidal radius as

$$r_\mathrm{t} = \left[\frac{Gm(<r_\mathrm{t})}{\Omega^2 - \frac{\mathrm{d}^2\Phi}{\mathrm{d}R^2}|_{\boldsymbol{R}}}\right]^{1/3} = \left[\frac{m(<r_\mathrm{t})/M(<R)}{2 + \frac{\Omega^2 R^3}{GM(<R)} - \frac{\mathrm{d}\ln M}{\mathrm{d}\ln R}|_R}\right]^{1/3}, \tag{1.6}$$

where the second equality is valid when the mass profile of the host is spherically symmetric. Here, $\boldsymbol{\Omega}$ is the instantaneous angular velocity of the subhalo with respect to the host center, $\boldsymbol{R}$ is the host-centric position of the subhalo, $m(<r)$ and $M(<R)$ are the enclosed mass profiles of the subhalo and host, and Φ is the gravitational potential of the host. In general, $r_\mathrm{t} \propto R$; hence, subhaloes that penetrate close to the host center during their pericentric passage are the most vulnerable to mass stripping. As we will demonstrate in this dissertation, a relatively simple model of tidal stripping, which employs the King (1962) prescription, can serve as the core component of a highly accurate model of subhalo mass loss.

The pericentric passage of a subhalo is a high-speed event, during which the fast change in the host potential drives a conversion of orbital energy into subhalo internal energy. The energy injected into the subhalo by *tidal heating* has been estimated via the impulse approximation (Spitzer, 1958) by Gnedin et al. (1999a) to be

$$\Delta E = \frac{4}{3}G^2 m\left(\frac{M}{v_\mathrm{p}}\right)^2 \frac{\langle r^2\rangle}{R_\mathrm{p}^4} f(R_\mathrm{p}) C_\mathrm{ad} \tag{1.7}$$

(e.g., Mo et al., 2010). Here, R_p is the radius of orbital pericenter (i.e., the impact parameter), v_p is the velocity of the subhalo at pericentric passage, $\langle r^2\rangle$ is the mass-weighted mean r^2 of the subhalo, $f(R_\mathrm{p})$ is a

factor that takes into account the extended mass distribution of the host (Gnedin et al., 1999b), and $C_{\rm ad}$ is an additional factor that takes into account the adiabatic shielding of particles near the subhalo center (Gnedin & Ostriker, 1999). In general, $f(R_{\rm p})$ ranges from unity when $R_{\rm p}$ is large to zero when $R_{\rm p}$ tends towards zero. When this is combined with the fact that $\Delta E \propto R_{\rm p}^{-4}$, it can be shown that ΔE attains a maximum when $R_{\rm p}$ is roughly the half-mass radius of the host. Due to the impulsive nature of the encounter, the energy added to the subhalo only changes the particle velocities but leaves the initial positions (and thus the subhalo potential) intact. Hence, after an impulsive heating event, the subhalo revirializes: the depth of its potential well decreases and it "puffs up" such that the central density decreases and material becomes distributed further out.

This energy injection can be put into context by comparing it to the internal binding energy of the subhalo prior to the encounter. For a NFW subhalo that is truncated at its virial radius, the binding energy is $E_{\rm b} = -\frac{1}{2}mV_{\rm v}^2 f_{\rm E}(c)$, where m is the subhalo virial mass, $V_{\rm v}$ is the circular velocity at its virial radius, and $f_{\rm E}(c)$ is a simple function of the halo concentration (Mo et al., 1998). Tidal heating can easily inject $\Delta E \sim |E_{\rm b}|$ into the subhalo during a pericentric passage. Indeed, van den Bosch et al. (2018) computed $\Delta E/|E_{\rm b}|$ for newly accreted subhaloes in the *Bolshoi* simulation, finding a median value of 1.9 and extreme ratios as large as 100. The naive assumption is that subhaloes that experience an injection of energy with $\Delta E \geq |E_{\rm b}|$ should be destroyed. However, ΔE only specifies the *total* energy injected into the subhalo but does not specify how the energy is distributed radially. Indeed, prior to integrating to get ΔE, it can be shown that $\langle \Delta E \rangle (r) \propto r^2$ — the particles in the outskirts of the subhalo receive the majority of the energy injection. Combining this with the impact of adiabatic shielding, one can conclude that NFW subhaloes should be nearly impossible to fully disrupt via tidal heating. As we discuss at length in chapter 4, the incorrect assumption that $\Delta E \sim |E_{\rm b}|$ implies subhalo disruption has been used to justify the results of numerical simulations in the past (e.g., Gonzalez-Casado et al., 1994; D'Onghia et al., 2010).

The previously described processes capture the evolution of the internal structure of the subhalo. Dynamical friction (DF), on the other hand, is responsible for mergers and the subsequent evolution of the subhalo *orbit*. In particular, DF causes subhaloes, globular clusters, and black holes alike to lose orbital angular momentum and ultimately sink to the centers of their host systems. The most frequently used DF formalism was introduced by Chandrasekhar (1943), which is based on the highly idealized case of a point mass (m) moving through an infinite uniform sea of background particles (with infinitesimal masses). Here, the background particles are deflected by the subject, forming a wake that imparts a drag force on the subject. While this is actually an incorrect description of the underlying cause of DF, the resulting model has proven quite successful at predicting the timescales of a range of astrophysical merger processes (for more sophisticated models of DF, see, e.g., Colpi et al., 1999; Banik & van den Bosch, 2021). By considering grav-

itational interactions between the subject and background particles as individual two-body encounters, Chandrasekhar summed over all such encounters in order to compute the net frictional force acting on the subject:

$$\boldsymbol{F}_{\rm DF} = -\frac{4\pi G^2 m^2}{v^2} \ln\Lambda\, \rho(<v)\, \frac{\boldsymbol{v}}{v}. \tag{1.8}$$

Here, $\boldsymbol{v}$ is the relative velocity of m with respect to the background, $\rho(<v)$ is the density of background particles with a speed below v, and $\ln\Lambda$ is the Coulomb logarithm. The background particles are often assumed to follow a Maxwell-Boltzmann distribution. A variety of forms have been used for the Coulomb logarithm, such as $\ln\Lambda = \ln(b_{\rm max}/b_{\rm min})$, with $b_{\rm max}$ and $b_{\rm min}$ the maximum and minimum impact parameters between the subject and the background particles that contribute to the DF drag. For the purpose of estimating the merger timescale of dark matter subhaloes in cosmological simulations, $\ln\Lambda = \ln(1 + M/m)$, where M is the mass of the host halo, has been shown to serve as a successful 'fudge factor' (Boylan-Kolchin et al., 2008; Jiang et al., 2008). Since $\boldsymbol{F}_{\rm DF} \propto m^2$, DF only has a significant impact on the orbits of massive subhaloes. Indeed, subhaloes with $m \lesssim 0.01M$ do not experience significant orbital decay over a Hubble time (e.g., Taffoni et al., 2003).

1.3.2 Artificial disruption

The history of dark matter substructure in cosmological simulations has been somewhat tumultuous. As discussed above, prior to the late 1990s (e.g., Ghigna et al., 1998), cosmological simulations did not have sufficiently high resolution to resolve subhaloes and suffered from numerical 'overmerging'. The underlying causes of overmerging have been attributed to particle–subhalo two-body heating (e.g., Carlberg, 1994; van Kampen, 1995) and inadequate softening of forces (e.g., Moore et al., 1996; Klypin et al., 1999a). However, as available computational power increased and cosmological simulations began to leverage higher resolutions than ever before, the problem of insufficient substructure turned into the opposite problem of too much substructure (e.g., Klypin et al., 1999b; Moore et al., 1999). This "missing satellite problem" was originally thought to be a serious issue for the viability of ΛCDM (for a review, see e.g., Bullock & Boylan-Kolchin, 2017) but has since been solved via developments in our understanding of the galaxy–halo connection (e.g., Kim et al., 2018).

Unfortunately, numerical overmerging has not yet been fully eradicated from high-resolution cosmological simulations and its impact is now much more subtle. The insufficient resolution of small subhaloes can still adversely impact the overall subhalo statistics, which introduces significant roadblocks for progress in small-scale clustering studies (e.g., Moster et al., 2018) and semi-analytical models of galaxy formation (e.g., Kitzbichler & White, 2008). In an effort to combat the impact of overmerging, simulation studies tend

to only consider subhaloes above a limit of $\sim 50 - 100$ particles. This particle limit has been estimated based on the convergence of subhalo mass functions across various simulations (e.g., Springel et al., 2008; Onions et al., 2012). Recently, van den Bosch (2017) performed a comprehensive study of the evolution pathways of subhaloes in the state-of-the-art *Bolshoi* simulation. The author found that subhaloes disrupt at a rate of roughly 13% per Gyr, which implies that ∼65% of subhaloes accreted at $z = 1$ will have been disrupted by the present day (Han et al., 2016; Jiang & van den Bosch, 2017). Furthermore, 20% of all subhalo disruption occurs above the 50 particle "convergence limit". These findings beg the question: what fraction of subhalo disruption seen in simulations is real, physical disruption and what fraction is still due to artificial disruption (i.e., overmerging)?

In van den Bosch et al. (2018), the authors set out to address this question. Using both analytical arguments and idealized N-body simulations, the authors assessed the impact of tidal stripping and heating (the two main disruptive processes) on NFW subhaloes. The found that NFW subhaloes can experience tidal shocks that are *many* multiples of their binding energy and still survive. Furthermore, they demonstrated that a subhalo will survive when a large fraction of its outer mass is instantaneously stripped, even when the remnant has positive binding energy. Thus, the authors concluded that physical disruption of CDM subhaloes should be exceedingly rare, consistent with previous analytical studies (e.g., Peñarrubia et al., 2010), with the implication being that the majority of disruption seen in cosmological simulations is likely to be artificial. In a follow-up study, van den Bosch & Ogiya (2018) ran a suite of idealized simulations of individual N-body NFW subhaloes orbiting within a static, analytical host halo potential in order to assess the conditions necessary for artificial disruption to take place. By varying the mass resolution (i.e., the number of particles that make up the initial subhalo) and the force softening length across the different simulations, the authors demonstrated that the bound remnant of a sufficiently resolved subhalo should survive perpetually (in the absence of baryonic physics). A large number of subhaloes in state-of-the-art cosmological simulations lie in the region of numerical parameter space that corresponds to being inadequately resolved. These subhaloes are subject to runaway instabilities driven by discreteness noise. Ultimately, van den Bosch & Ogiya (2018) doubled down on their previous conclusion: the majority of subhalo disruption in simulations must be artificial. As we discuss below, this puts considerable limitation on the utility of cosmological simulations and also adversely impacts various simulation-calibrated semi-analytical models.

1.3.3 Modeling approaches

Due to the non-linear nature of hierarchical structure formation, a fully analytical prescription of the build-up and subsequent evolution of substructure has proven to be intractable. Because of this, cosmological

N-body simulations have been the primary method of studying subhalo populations. Unfortunately, as we discuss above, these simulations still suffer from numerical limitations that can impact the predicted subhalo statistics. This can be mitigated by increasing the simulation resolution, which comes at the cost of increasing the computational demands and greatly reducing the sample size of host systems available for analysis. Semi-analytical modeling is an alternative approach that attempts to leverage various effective models of subhalo evolution processes (Section 1.3.1) in order to predict subhalo statistics in a computationally efficient manner while yielding much larger host halo sample sizes and circumventing some of the numerical limitations of N-body simulations (such as the mass resolution of individual subhaloes). Much like simulations, these semi-analytical models are capable of generating "subhalo catalogs", which provide time series of the mass, position, velocity, and various internal properties for each subhalo as it evolves within its host. A wide range of semi-analytical models have been presented in the literature (e.g., Taylor & Babul, 2001; Peñarrubia & Benson, 2005; van den Bosch et al., 2005b; Zentner et al., 2005a; Kampakoglou & Benson, 2007; Pullen et al., 2014; Jiang & van den Bosch, 2016), the majority of which share a core set of components:

- Analytical merger trees that are generated using EPS theory-based algorithms (e.g., Cole et al., 2000; Parkinson et al., 2008; Benson, 2017).
- A model that describes the distribution of orbital properties of infalling subhaloes, which can be used to sample initial orbits (e.g., Jiang et al., 2015; Li et al., 2020).
- An orbit integrator, which includes a prescription for dynamical friction (e.g., Chandrasekhar, 1943).
- Models of the density structure of the host and (initial) subhaloes (e.g., Navarro et al., 1997).
- A model that captures the response of the subhalo internal structure to tidal heating and stripping (e.g., Hayashi et al., 2003; Peñarrubia et al., 2008).
- A prescription for subhalo mass-loss (e.g., Zentner et al., 2005a).

Several of the components in a semi-analytical model have free parameters that are calibrated such that the predicted subhalo statistics are in good agreement with cosmological simulations. In particular, the mass-loss rate is tuned to reproduce the subhalo mass function and the merger tree algorithm is fit to correctly capture the conditional mass function. This procedure introduces an obvious limitation — any systematic biases in the simulation results will propagate through to the semi-analytical model results via the calibration. In particular, if simulations still suffer from a significant amount of artificial subhalo disruption, this will adversely bias, e.g., the inferred mass-loss rate. These problems were hinted at in Jiang (2016), which prompted the subsequent development of a new, modular semi-analytical modeling frame-

work (`SatGen`) that ultimately serves as the cornerstone of this dissertation (Jiang et al., 2021). As we will show, a significant part of our work is aimed at calibrating this model such that its results are not adversely impacted by any potential numerical overmerging that may still impact cosmological simulations. Because of the simulation-independent nature of this model, we are also able to use it to provide a complementary assessment of the state of artificial disruption in such simulations.

1.4 Observational constraints

The population of dark matter subhaloes is intimately connected to the underlying particle nature of dark matter. As discussed above, CDM models predict a SHMF that follows a power law down to Earth masses ($\sim 10^{-6}\,M_\odot$). On the other hand, WDM results in a suppression of low-mass subhaloes (with $m \lesssim 10^9\,M_\odot$) due to its larger free-streaming velocity (e.g., Lovell et al., 2014; Bose et al., 2017). However, substructure is not only sensitive to the dark matter temperature. For example, a minimal dark sector that consists of a massive particle and a "dark photon" would allow for self-interactions, which would reduce the inner halo density slope to create a "core" (e.g., Kaplinghat et al., 2016). Unlike NFW haloes, cored haloes are less resilient to the impact of tides (Peñarrubia et al., 2010); hence, a self-interacting dark matter (SIDM) particle would result in an overall reduction in subhalo abundance. If the particle was instead an ultra-light boson (i.e., fuzzy dark matter [FDM]), the halo center would consist of a Bose–Einstein condensate "soliton" with a radius of order kpc; in this case, the halo center would also be cored, ultimately suppressing substructure (e.g., Robles et al., 2017; Burkert, 2020). Furthermore, the formation history and abundance of haloes and subhaloes alike is sensitive to the cosmology, most directly the matter energy density (i.e., Ω_m).

Clearly, any observational constraints placed on the abundance and structure of dark matter subhaloes would map to constraints on the particle nature of dark matter and/or the underlying cosmology. Hence, a variety of observational probes are currently being employed to search for signatures of substructure. Since (satellite) galaxies are expected to form and reside within dark matter (sub)haloes (e.g., White & Rees, 1978), there exists a direct link between the observed abundance of galaxies and the inferred abundance of haloes. The "galaxy–halo connection" is typically codified via the halo occupation distribution (HOD; e.g., Zheng et al., 2005) or the conditional luminosity function (CLF; e.g., van den Bosch et al., 2013), models which are calibrated by comparing mock galaxy catalogs (generated by populating haloes from cosmological N-body simulations with galaxies) to samples from large-scale surveys. These models assume that the galaxy luminosity is only a function of halo mass, which is an oversimplification due to *assembly bias*. For example, galaxies are also sensitive to the halo formation history (e.g., Zentner et al., 2014; Mao et al., 2015) and environment (e.g., Croton et al., 2007). Recently, however, sophisticated forward modeling-based approaches

to Milky Way satellite galaxy abundance matching have been used to place some of the tightest constraints to date on non-CDM dark matter models (Nadler et al., 2021).

Gaps identified in stellar streams have begun to stand out as potential smoking-gun indicators of the presence of dark matter subhaloes (e.g., Erkal et al., 2016; Banik et al., 2019; Bonaca et al., 2020). For example, Bonaca et al. (2020) used high-resolution spectroscopy of stars in a spur of the GD-1 stellar stream to localize the orbit of a perturber, which could be either a globular cluster or a compact subhalo. On the theoretical side, Erkal et al. (2016) modeled encounters between stellar streams and CDM subhaloes in order to predict the expected number and size of gaps, concluding that the current quantity of observed gaps is consistent with ΛCDM predictions. However, future stream studies, made possible by *Gaia* (Gaia Collaboration et al., 2016), should be sensitive enough to place much tighter constraints.

Substructure can also be probed via gravitational lensing distortions. Subhaloes in the lens halo can significantly perturb the magnification pattern of the source, which depends on the second derivative of the lens potential. Spatially extended background sources form arcs in the lensed image. Subhaloes near the arc distort the lensed image relative to that predicted for the case of a smooth lens, which enables this approach to both localize the perturbing subhalo and estimate its mass (down to masses of $10^8 - 10^9\ M_\odot$ Vegetti et al., 2014; Hezaveh et al., 2016). Multiply lensed point sources (e.g., quasars) yield different observables, namely the image positions and magnification ratios between the images. These "flux ratios" are highly sensitive to the impact of subhaloes on the second derivative of the lens potential, enabling the localization and mass estimate of subhaloes down to $\sim 10^7\ M_\odot$. Recently, Gilman et al. (2020a) used eight quadruple-image lenses to place tight upper bounds on the free-streaming mass (and thus the dark matter particle mass), greatly disfavoring WDM models.

The prevailing CDM particle model is the WIMP, which is searched for via both direct and indirect detection experiments. Direct approaches involve placing sensitive detectors deep underground and waiting for passing dark matter particles to undergo scattering events with the detector material (e.g., Schumann, 2019). Indirect approaches (e.g., Stref & Lavalle, 2017; Somalwar et al., 2021), on the other hand, attempt to search for γ-ray signals of particle annihilation or decay in astrophysical sources that are expected to be dark matter-rich (e.g., the center of the Milky Way or its satellite galaxies). Because annihilation is a two-body process, its signal is proportional to the square of the dark matter density. Hence, the presence of compact subhaloes can greatly boost the expected annihilation signal. The total annihilation flux is directly proportional to the J-factor, which captures the spatial distribution of dark matter along the line of sight as

$$J = \int_\Omega \int_{\mathrm{los}} \rho^2[r(\theta, \mathcal{D}, l)]\,\mathrm{d}l\,\mathrm{d}\Omega. \tag{1.9}$$

Here, ρ is the density profile of the observed halo, Ω is the solid angle, and the radius from the center of the halo, r, depends on the angle of observation, θ, and distance to the halo along the line of sight, $\mathcal{D}$. The total J-factor of a host system of mass M that contains subhaloes is *boosted* to $J(M) = [1 + B(M, m_{\rm cut})]\tilde{J}(M)$, where $\tilde{J}(M)$ is the contribution from the smooth host halo. Hence, the contribution from subhaloes down to the cut-off mass, $m_{\rm cut}$, is

$$B(M, m_{\rm cut})\tilde{J}(M) = \int \frac{{\rm d}N}{{\rm d}m} J(m) {\rm d}m, \tag{1.10}$$

where ${\rm d}N/{\rm d}m$ is the subhalo mass function. We emphasize that this is the total contribution by subhaloes taken over an entire host halo. In reality, the subhalo contribution itself is a function of the distance of the observation from the host center since the subhalo distribution is radially dependent. The boost factor can be considerably larger than unity, indicating that subhaloes contribute more to the total annihilation signal than the smooth host halo background itself. However, $B(M, m_{\rm cut})$ is highly uncertain — it is extremely sensitive to the properties of subhaloes below the resolution limit of cosmological simulations and varies greatly with $m_{\rm cut}$, which depends on the dark matter particle model. For example, Strigari et al. (2007) finds that, for a $M = 10^8\,M_\odot$ host halo (corresponding to a dwarf spheroidal galaxy), $B(M) \lesssim 41$ if $m_{\rm cut} = 10^{-13}\,M_\odot$ but $B(M) \lesssim 2$ if $m_{\rm cut} = 10^{-2}\,M_\odot$. Clearly, due to the limitations of simulations, semi-analytical models of substructure must play a considerable role in making accurate model predictions for CDM J-factors and boost factors (e.g., Bartels & Ando, 2015) in order to use γ-ray observations to successfully constrain dark matter properties (primarily the WIMP mass and annihilation cross section).

As this wide variety of subhalo-sensitive observational techniques matures, it is becoming increasingly important that predictions of subhalo abundance and structure are accurate and true representations of the underlying model (be it CDM, WDM, SIDM, or FDM). These predictions must include robust estimations of the halo-to-halo variance in quantities of interest in order to be useful for inference. Thus, predictive models must be able to circumvent the current limitations of simulation-based approaches, most crucially the mass resolution limit and small host halo sample sizes.

1.5 Structure of this dissertation

In this dissertation, we build an improved semi-analytical model of dark matter substructure evolution and use it to assess various limitations of current-generation cosmological N-body simulations. This work begins in chapter 2 (originally published in Green & van den Bosch, 2019), where we develop a state-of-the-art model of the tidally evolved subhalo density profile (ESHDP), which captures the joint effects of tidal heating and stripping on the subhalo structure. We calibrate this model using a large suite of high-

resolution idealized N-body simulations of individual subhaloes orbiting within a static, analytical host halo potential that are free from the influence of artificial disruption (the *DASH* simulations, described in Appendix A and introduced in Ogiya et al., 2019).

Following our initial release of the `SatGen` semi-analytical modeling framework (described in Appendix B and introduced in Jiang et al., 2021), we set out to improve it in several key ways. In chapter 3 (originally published in Green et al., 2021a), we use our ESHDP model to build a prescription for the subhalo mass-loss rate that is calibrated to reproduce the mass evolution of *DASH* subhaloes. We then augment `SatGen` by incorporating into it our ESHDP and mass-loss models, improving the dynamical friction routine, and introducing a sophisticated model of the orbits of infalling subhaloes. By cataloguing all instances of subhalo disruption in the *Bolshoi* simulation (Klypin et al., 2011), we construct a probabilistic framework for artificial disruption that can be applied to `SatGen` results in order to emulate a simulation. Using the "disruption-free" semi-analytical model, we generate ensembles of subhalo populations for host haloes of a range of masses. We compare subhalo mass functions, radial profiles, and substructure mass fractions between the model and *Bolshoi*, demonstrating that when we apply the probabilistic model of artificial disruption, we are able to reproduce the simulation results with exquisite accuracy. We place estimates on the overall impact of artificial disruption on simulated subhalo populations and argue that `SatGen` provides a promising complementary approach to N-body simulations for future subhalo studies.

Having demonstrated several powerful use cases for `SatGen` in Jiang et al. (2021) and Green et al. (2021a), we continue forward in chapter 4 (originally published in Green et al., 2021b) by applying the semi-analytical model to assess the impact of a central galactic disc potential on subhalo populations. Using a set of 10,000 merger trees for Milky Way-mass host haloes, we generate evolved subhalo populations using a variety of composite halo–disc host potentials. The mean impact of the disc is to suppress the overall abundance of subhaloes; the strength of this suppression grows as host-centric radius decreases. The level of suppression that we report is in agreement with prior cosmological and idealized simulation studies. However, our study is the first of its kind to have access to such a large host halo sample. Because of this, we are able to demonstrate that the halo-to-halo variance of most substructure statistics (i.e., mass functions and radial profiles) is far larger than the mean impact on such statistics due to the presence of a disc. By re-evolving identical trees using different host potentials, we are able to study the differential influence of a disc on the $z = 0$ mass of *individual* subhaloes. Specifically, we show that the disc drives excess subhalo mass loss, the strength of which increases with decreasing pericentric radius. We argue against the notion that a disc causes subhalo disruption via disc shocking (first promoted by D'Onghia et al., 2010) and instead use our results to demonstrate that the chief effect of the disc is simply to increase the central density concentration of the host, which ultimately promotes excess mass loss via tidal stripping.

The studies described above constitute the primary aim of this dissertation, which is focused on the evolution of dark matter subhaloes. However, after demonstrating in chapter 3 and chapter 4 the significance of halo-to-halo variance for properly interpreting substructure statistics, a slightly broader theme emerges. The influence of variance in halo mass accretion histories has astrophysical and cosmological importance that extends beyond subhalo studies. For example, in chapter 5 (originally published in Green et al., 2020), we employ a Monte Carlo-based mass accretion history generator (also used in `SatGen`) in combination with an analytical model of non-thermal pressure production to generate Sunyaev-Zel'dovich effect (SZ; Sunyaev & Zeldovich, 1972) signal observations for mock galaxy clusters. We demonstrate that a substantial fraction of the scatter in the halo mass–SZ signal relation that is measured from cosmological simulations can be reproduced by our simple model, thus concluding that variance in cluster mass assembly histories is a primary culprit of such scatter.

In chapter 6 (originally published in Green et al., 2019), we continue our focus on galaxy clusters by exploring novel approaches to estimating the halo mass from observations. Specifically, we employ a simple machine learning algorithm (random forest regression) to predict the underlying halo mass given a range of summary statistics computed from mock X-ray observations of simulated clusters. We find that the most informative statistics (aside from the *most* important one, the bolometric luminosity) for mass prediction include several "morphological parameters", such as the surface brightness concentration, smoothness, and asymmetry. These parameters encode important information about the mass accretion history of the cluster, providing a physical basis for their utility in mass estimation. We demonstrate that a random forest model, properly armed with such morphological parameters, can offer a modest improvement in mass estimation accuracy relative to simple linear regression techniques typically used in observational studies.

We conclude this dissertation by summarizing our key results from each study in chapter 7 and discussing potential future `SatGen`-based studies in chapter 8.

Chapter 2

Density profile evolution

This chapter has been published as an article by Green & van den Bosch (2019) in the Monthly Notices of the Royal Astronomical Society by Oxford University Press.

2.1 Background

In the Λ cold dark matter (ΛCDM) cosmological model of structure formation, primordial density perturbations with a scale-invariant power spectrum collapse to form virialized haloes. Due to the negligible free-streaming velocities of CDM, haloes form on all scales, with smaller perturbations collapsing earlier and subsequently assembling from the bottom up to form more massive haloes. Since 1997, cosmological N-body simulations have shown that the dense, inner regions of these smaller haloes continue to live on as subhaloes within their hosts after having been accreted (Tormen et al., 1997; Moore et al., 1998; Ghigna et al., 1998), and these subhaloes themselves host sub-subhaloes, and so on, forming a complete hierarchy of substructure (Gao et al., 2004; Springel et al., 2008; Giocoli et al., 2010). As these subhaloes orbit their hosts, they are subjected to various forces that work to disrupt them, including dynamical friction, tidal stripping and impulsive heating due to the host, and harassment by other substructure (e.g., Mo et al., 2010; van den Bosch et al., 2018).

The statistics of dark matter (DM) substructure are sensitive to the underlying DM model. In particular, the DM thermal velocity sets the cutoff scale for low-mass haloes, which in turn impacts the abundance of substructure (e.g., Knebe et al., 2008; Lovell et al., 2014; Colín et al., 2015; Bose et al., 2017), and the (potentially nonzero) cross-section for DM self-interaction can core out the otherwise cuspy slopes of subhalo inner density profiles, making them less resilient to the strong tidal forces of the host halo (e.g., Burkert, 2000; Vogelsberger et al., 2012; Rocha et al., 2013). The primary observational techniques used to

probe the properties of DM substructure include gravitational lensing (e.g., Dalal & Kochanek, 2002; Keeton & Moustakas, 2009; Vegetti et al., 2014; Hezaveh et al., 2016; Gilman et al., 2020a), gaps in stellar streams (e.g., Carlberg, 2012; Ngan & Carlberg, 2014; Erkal et al., 2016), and indirect detection via DM annihilation and decay signals (e.g., Strigari et al., 2007; Pieri et al., 2008; Hayashi et al., 2016; Hiroshima et al., 2018; Delos, 2019). Furthermore, since satellite galaxies are expected to reside within some fraction of the DM subhaloes, the demographics of DM substructure has a direct correspondence to that of satellite galaxies (e.g., Vale & Ostriker, 2006; Hearin et al., 2013; Behroozi et al., 2013b; Newton et al., 2018), which ultimately impacts small-scale clustering statistics (e.g., Benson et al., 2001; Berlind et al., 2003; Kravtsov et al., 2004; Campbell et al., 2018). Thus, being able to accurately predict the abundance and structural evolution of DM subhaloes is paramount for using astrophysics to study the particle nature of dark matter.

Due to its high nonlinearity, a purely analytical description of subhalo evolution is impossible, even in the most idealized of circumstances (for a detailed discussion, see van den Bosch et al., 2018). Hence, the primary method employed for studying the demographics of DM substructure has been, and remains, cosmological N-body simulations. Prior to the late 1990s, numerical simulations did not yet have sufficient mass and force resolution to resolve surviving populations of subhaloes (Moore et al., 1996; Klypin et al., 1999a). As increased computational power has enabled access to ever higher resolutions, many convergence tests have since been performed to validate the results of more recent N-body simulations, demonstrating consistent subhalo mass functions above a resolution limit of 50-100 particles (e.g., Springel et al., 2008; Onions et al., 2012; Knebe et al., 2013; van den Bosch & Jiang, 2016; Griffen et al., 2016); however, mass function convergence is only a necessary, but not sufficient, condition to guarantee the physical correctness of numerical simulations. Van den Bosch (2017) showed that the complete disruption of subhaloes occurs very frequently in state-of-the-art simulations, with a mass function of disrupted subhaloes that is identical to that of the surviving population. The inferred disruption rate implies that roughly 65% of subhaloes accreted around $z = 1$ are disrupted by $z = 0$ (Han et al., 2016; Jiang & van den Bosch, 2017). Some authors have argued that complete disruption is a physical consequence of tidal heating and/or tidal stripping (Hayashi et al., 2003; Taylor & Babul, 2004; Klypin et al., 2015). However, van den Bosch et al. (2018) demonstrated that neither tidal heating nor tidal stripping are independently sufficient to completely disrupt CDM subhaloes, a result consistent with the idealized, high-resolution numerical simulations of Peñarrubia et al. (2010). Van den Bosch & Ogiya (2018) ran a suite of similar, idealized numerical experiments, finding that subhalo disruption in N-body simulations is largely due to two key numerical details: (i) discreteness noise due to insufficient particle resolution and (ii) inadequate force softening. The optimal force softening criteria put forth by van den Bosch & Ogiya (2018) have since been corroborated by Ludlow et al. (2019) and are in good agreement with the criteria of Zhang et al. (2019).

This artificial subhalo disruption may have substantial consequences across cosmology and astrophysics. For example, in small-scale clustering analysis, the uncertainty due to disruption reduces the predictive power of methods such as subhalo abundance matching (e.g., Vale & Ostriker, 2006; Conroy et al., 2006; Guo et al., 2010; Hearin et al., 2013), while the reduced abundance of substructure implies that dark matter annihilation boost factors (e.g., Bergström et al., 1999; Ando et al., 2019) may be substantially underestimated. The all-important, outstanding question is to what extent this artificial disruption impacts the subhalo mass and/or velocity function predicted by cosmological simulations. The work of van den Bosch & Ogiya (2018) suggests that the answer is unlikely to come from numerical simulations, as there is no obvious way to circumvent the numerical issues. Instead, we may hope to gain some insight from semi-analytical models of the build-up and evolution of dark matter substructure (e.g., Taylor & Babul, 2001; Peñarrubia & Benson, 2005; Zentner et al., 2005a; van den Bosch et al., 2005b; Kampakoglou & Benson, 2007; Gan et al., 2010; Pullen et al., 2014). The problem, though, is that the lack of a complete theory of tidal evolution implies that these semi-analytical models need to be calibrated, which is typically done by tuning the model to reproduce the subhalo mass functions inferred from cosmological N-body simulations. This obviously implies that the models inherit the shortcomings of the simulations. The main goal of this chapter is to present a model of the evolution of subhalo density profiles that circumvents this catch-22 situation.

Before describing our methodology, though, it is insightful to try to estimate how big of an impact artificial disruption may potentially have. We can do so using the semi-analytical model of Jiang & van den Bosch (2016), which combines halo merger trees with simple models of the tidal evolution of subhaloes, to predict the evolved subhalo mass and velocity functions of dark matter substructure (see Jiang & van den Bosch, 2017). The model treats both mass stripping as well as subhalo disruption, the efficiencies of which are calibrated to reproduce the results of the high-resolution `Bolshoi` simulation (Klypin et al., 2011). The left- and right-hand panels of Fig. 2.1 plot the subhalo mass and velocity functions, respectively. The solid circles indicate the results from the `Bolshoi` simulation for present-day host haloes with masses in the range $14.0 \leq \log[M_{\rm h}/\,h^{-1}{\rm M}_\odot] \leq 14.5$, while the solid line is the model prediction from Jiang & van den Bosch (2016). Since the latter is calibrated against the former, it should not come as a surprise that the model fits the simulation data well. However, as discussed at length in Jiang & van den Bosch (2016), crucial for this success is the separate treatment of subhalo disruption. We can now use this model to predict what the subhalo mass and velocity functions would look like under the assumption that *all* disruption is artificial. To that extent, we rerun the same model, this time turning off disruption; in this case, subhaloes continue to experience mass loss rather than fully disrupt. The resulting mass and velocity functions are indicated by the dashed curves. Clearly, artificial disruption does not merely impact the mass/velocity functions at the low mass end, close to the resolution limit of the simulation; rather, the mass and velocity functions are

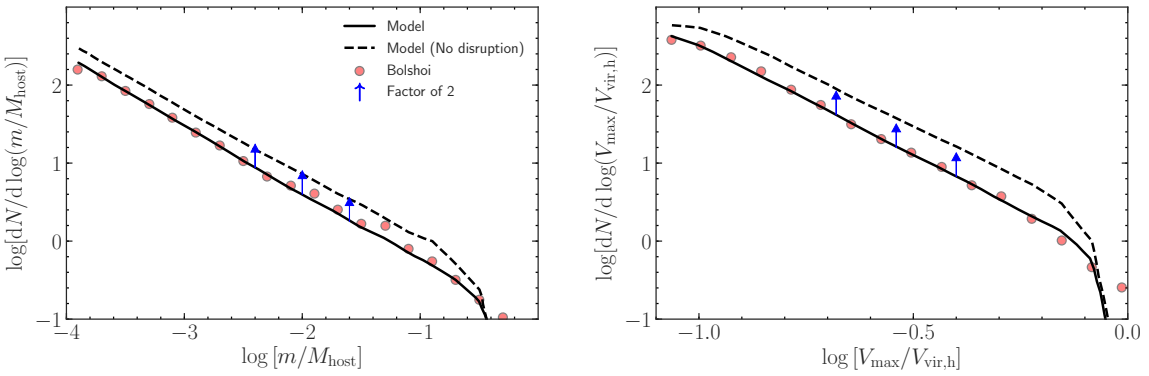

Figure 2.1: Subhalo mass (*left*) and velocity (*right*) functions for host haloes with masses in the range $14.0 \leq \log[M_{\rm h}/\,h^{-1}{\rm M}_\odot] \leq 14.5$. Symbols indicate the results obtained from the `Bolshoi` simulation (Klypin et al., 2011), whereas the solid lines are the results obtained from the semi-analytical model of Jiang & van den Bosch (2016). The latter includes models for subhalo mass loss and subhalo disruption that have been tuned to specifically reproduce the subhalo mass and velocity functions of the `Bolshoi` simulation. The dashed lines show the predictions of the same model, but with subhalo disruption turned off. The inference is that if the majority of subhalo disruption is artificial, as claimed by several recent studies (Peñarrubia et al., 2010; van den Bosch et al., 2018; van den Bosch & Ogiya, 2018), state-of-the-art cosmological simulations may under-predict the abundance of subhaloes by as much as a factor of two (blue arrows). See text for a more detailed discussion.

boosted globally by factors of ~ 2 and ~ 2.5, respectively. If these admittedly crude predictions are even remotely correct, the implications are far-reaching. It suggests that state-of-the-art cosmological simulations systematically under-predict the abundance of substructure by as much as a factor of two, which, interestingly, is precisely what is needed to solve the 'galaxy clustering crisis' in subhalo abundance matching (Campbell et al., 2018). At the very least, these results signal the need to carefully examine the tidal evolution of subhaloes in more detail, which is the core-motivation behind the study presented here.

Semi-analytical models of the build-up and evolution of dark matter substructure consist of three main ingredients: (i) a halo merger tree, which quantifies the subhalo masses and redshifts at accretion, (ii) a model of the orbital evolution, including dynamical friction and self-friction (Miller et al., 2020), and (iii) a model that describes how the mass and density profile of a subhalo evolves subject to the tidal forces that it experiences. Semi-analytical merger tree algorithms are calibrated using merger histories from cosmological simulations, which depend on the halo properties at infall and are therefore less sensitive to the effects of artificial disruption than the evolution of individual subhaloes. On the other hand, the evolution of the subhalo density profile typically requires a model of how the bound mass of the subhalo evolves with time and how this affects the subhalo's density profile. Neither of these can be treated analytically from first principles, and the models therefore typically rely on parametrized treatments that somehow need to be calibrated. In order to prevent the catch-22 situation eluded to above, in Ogiya et al. (2019) we introduced the Dynamical Aspects of SubHaloes (DASH) database, a large library of idealized, high-resolution N-body

simulations of the tidal evolution of individual subhaloes. These simulations cover a wide range of relevant subhalo parameters (i.e., orbital energy and angular momentum at infall and halo concentrations) and are evolved with sufficient numerical resolution to assuage the impact of discreteness noise and insufficient force softening. As a next step towards building a more accurate semi-analytical treatment of dark matter substructure evolution, the present chapter sets out to develop a new model of the tidal evolution of the subhalo density profile, calibrated against DASH and therefore unimpeded by numerical artifacts, that is applicable to a far wider range of subhalo parameter space than that of previous works (Hayashi et al., 2003; Peñarrubia et al., 2010; Drakos et al., 2017).

This chapter is organized as follows: Section 2.2 provides an overview of the DASH simulation database. In Section 2.3, we describe the methods used for building and calibrating our model of the evolved subhalo density profile and then quantify the model's capability of reproducing simulated subhalo density profiles. In Section 2.4, we demonstrate the model's performance at capturing the evolution of the subhalo structural parameters, $V_{\rm max}$ and $r_{\rm max}$. Lastly, in Section 2.5, we summarize the results and discuss future work.

2.2 The *DASH* database

The DASH library[1] (Ogiya et al., 2019) is a suite of idealized, collisionless N-body simulations that follow the evolution of an individual N-body subhalo as it orbits within the fixed, analytical potential of its host halo. Both the fixed host halo and the initial subhalo are spherically symmetric, each with a Navarro-Frenk-White (NFW; Navarro et al., 1997) density profile:

$$\rho_{\rm NFW}(r) = \rho_0 \left(\frac{r}{r_{\rm s}}\right)^{-1} \left(1+\frac{r}{r_{\rm s}}\right)^{-2}, \tag{2.1}$$

where the model parameters $r_{\rm s}$ and ρ_0 are the characteristic scale radius and density, respectively. The halo virial radius $r_{\rm vir}$ is defined to be the radius within which the average density is $\Delta_{\rm vir} = 200$ times the critical density of the Universe $\rho_{\rm crit}$. The corresponding virial mass is defined as $M_{\rm vir} = \frac{4\pi}{3}\Delta_{\rm vir}\rho_{\rm crit}r_{\rm vir}^3$. The halo concentration is defined as $c \equiv r_{\rm vir}/r_{\rm s}$, and the virial velocity as $V_{\rm vir} \equiv \sqrt{GM_{\rm vir}/r_{\rm vir}}$. Throughout this work, the subscripts 'h' and 's' represent quantities associated with the host- and subhaloes, respectively.

The initial conditions are generated assuming that the NFW subhalo has an isotropic velocity distribution, such that the phase-space distribution function (DF) depends only on energy. The simulations are performed with a tree code (Barnes & Hut, 1986) developed for graphics processing unit (GPU) clusters (Ogiya et al., 2013). Each subhalo is initially made up of 1,048,576 particles, forces are softened with a Plummer

1. https://cosmo.oca.eu/dash/

equivalent length $\epsilon = 0.0003 r_{\rm vir,s}$, and the opening angle of the tree is set to $\theta = 0.7$. Orbits are integrated with the second-order leapfrog scheme with a global, adaptive time step $\Delta t = \sqrt{\epsilon / a_{\rm max}}$, with $a_{\rm max}$ the maximum, absolute acceleration among all particles at that time. As demonstrated in van den Bosch & Ogiya (2018), these parameters are sufficient to properly resolve the subhalo evolution.

For each simulation, the library contains various data about the subhalo evolution at 301 snapshots, with a physical time interval between each of 0.12 Gyr. This corresponds to a total evolution time of 36 Gyr, or 2.5 to 12 radial periods depending on the orbital configuration. The subhalo is initially placed at the apocenter of its orbit. At each timestep, DASH contains the radial profiles of the subhalo density, enclosed mass, and radial/tangential velocity dispersion, as well as its bulk position, velocity, bound mass fraction $f_{\rm b}(t)$, and half-mass radius $r_h(t)$ (see Appendix A of van den Bosch et al. (2018) for details on how these quantities are computed). The radial profiles are computed for 40 logarithmically-spaced radial bins, which span $-2.95 \leq \log(r/r_{\rm vir,s}) \leq 0.95$. While all DASH simulations initially meet the numerical reliability criteria of van den Bosch & Ogiya (2018), the simulations can become unreliable as the bound mass fraction becomes small. In this work, we only consider simulation snapshots that meet the following two reliability criteria, introduced in van den Bosch & Ogiya (2018), each of which can be computed using $r_{\rm h}(t)$ and $f_{\rm b}(t)$. The first criterion, motivated by Power et al. (2003), demands that the softening length be sufficiently small to resolve the maximum particle accelerations, a requirement given by

$$f_{\rm b}(t) > 1.79 \frac{c_{\rm s}^2}{f(c_{\rm s})} \left(\frac{\epsilon}{r_{\rm vir,s}} \right) \left(\frac{r_{\rm h}(t)}{r_{\rm vir,s}} \right). \tag{2.2}$$

The second criterion, related to discreteness noise, states that the number of bound particles in the subhalo must exceed $N_{\rm crit} = 80 N^{0.2}$, with N the initial number of particles in the subhalo. Once the bound particle count falls below this value, the subhalo experiences a discreteness-driven runaway instability resulting in artificial disruption. In the DASH database, this requirement translates to

$$f_{\rm b}(t) = 1.22 \times 10^{-3}. \tag{2.3}$$

We note that over 99.5% of the DASH simulation snapshots meet the requirements of equations (2.2) and (2.3).

In addition to excluding snapshots that do not meet the numerical reliability criteria, we also perform several additional preprocessing steps. We exclude snapshots that are within the 10% of the orbital period centered around pericentric passage in order to avoid intervals where $f_{\rm b}(t)$ is changing rapidly and the

boundedness designation of individual particles is less reliable.[2] Additionally, only subhalo radial density profile points in the range $0.01 \leq r/r_{\rm vir,s} \leq 1$ are used for analysis; this innermost radius corresponds to ~3 times the softening length, inside of which the density profile is not reliable.

The database contains 2,253 simulations of subhaloes orbiting within host haloes with an initial host-to-subhalo mass ratio of $M_{\rm vir,h}/M_{\rm vir,s} = 1000$, a ratio sufficiently large that the effects of dynamical friction (Chandrasekhar, 1943) can safely be neglected. Furthermore, due to the self-similar nature of subhalo evolution, the simulations apply generally to initial configurations with $M_{\rm vir,h}/M_{\rm vir,s} \gtrsim 100$, regardless of the absolute value of $M_{\rm vir,h}$. The simulations spread a four-dimensional parameter space of host- and subhalo concentrations and initial orbital configurations, as illustrated by Figs. 2 and 4 in Ogiya et al. (2019). The concentrations $c_{\rm h}$ and $c_{\rm s}$ cover the range $3.1 \leq c \leq 31.5$, with the majority of the simulations devoted to the host- and subhalo concentrations (and ratios between the two) most commonly seen in cosmological simulations for haloes roughly in the range of $10^7 < M_{\rm vir}/(h^{-1}M_\odot) < 10^{15}$, determined using the method described in Section 2.2.3 of Ogiya et al. (2019). The initial orbital configuration is parametrized by two dimensionless analogs to energy and angular momentum: $x_{\rm c} \equiv r_{\rm c}(E)/r_{\rm vir,h}$, where $r_{\rm c}(E)$ is the radius of the circular orbit of energy E, and the circularity $\eta = L/L_{\rm c}(E)$, where L is the initial orbital angular momentum and $L_{\rm c}(E)$ is the angular momentum of the corresponding circular orbit with the same energy. The orbital parameters are sampled in the range $0 \leq \eta \leq 1$ (linearly) and $0.5 \leq x_{\rm c} \leq 2$ (logarithmically). The majority of the simulations are devoted to orbital parameters near the peak of the probability distribution seen at infall in cosmological simulations (Jiang et al., 2015).

2.3 Evolved subhalo density profile

The objective of this chapter is to calibrate a model of the evolution of the subhalo density profile against the DASH simulations. As described above, the DASH database consists of 2,253 simulations, each of which has 301 snapshots of time evolution over several orbital periods. At each of these snapshots, various radial profiles and global subhalo properties are stored. After performing the preprocessing steps described previously, the calibration dataset consists of a total of roughly 6×10^5 snapshots of subhalo evolution labeled by (i) the initial configurations, which span the parameter space of $c_{\rm h}$, $c_{\rm s}$, $x_{\rm c}$, and η values, and by (ii) the bound fractions $f_{\rm b}(t)$, which span roughly three orders of magnitude (~10^{-3} to 1). At each of these snapshots, we compute the ratio of the evolved subhalo density profile relative to the initial subhalo density profile, which we refer to as the *transfer function* $H(r,t) = \rho(r,t)/\rho(r,t=0)$, where $\rho(r,t=0)$ is the NFW

2. When this selection criterion is removed, our results remain qualitatively the same and we find that the variance in the residuals between our best-fit model and the DASH density profiles (as in Fig. 2.4) increases slightly at large subhalo radii.

profile of equation (2.1). The transfer function is stored for 20 radial bins spanning $0.01 \leq r/r_{\rm vir,s} \leq 1$ at each snapshot. This calibration dataset is immense, including over 10 million distinct data points of subhalo transfer functions.

The studies of Hayashi et al. (2003, hereafter H03) and Peñarrubia et al. (2008) argued that the subhalo density profiles depend solely on the density profile at infall and the total amount of mass lost thereafter. In particular, H03 describes the evolved density profile in terms of a transfer function, $H_{\rm H03}(r|f_{\rm b})$, which implies that the density profiles of subhaloes are insensitive to how and when they have lost their mass. Based on the same principle, Peñarrubia et al. (2010, hereafter P10) provides a prescription to obtain a transfer function based off of their "tidal track" fitting function for the structural parameters normalized by their initial values, $\frac{V_{\rm max}}{V_{\rm max,i}}(f_{\rm b})$ and $\frac{r_{\rm max}}{r_{\rm max,i}}(f_{\rm b})$. Here $V_{\rm max}$ is the maximum circular velocity and $r_{\rm max}$ is the associated radius. Based on the DASH database, though, we find that the residuals between these models and the DASH transfer functions exhibit a significant, systematic correlation with the initial subhalo concentration, $c_{\rm s}$. Neither H03 nor P10 observed this dependence, as both works only considered subhaloes with a single value for the concentration ($c_{\rm s} = 10$ and 23.1, respectively). In addition, we find that the dependence on $c_{\rm s}$ is much stronger than on any of $c_{\rm h}$, $x_{\rm c}$, or η, which illustrates that while the evolved subhalo density profile depends on both the total amount of mass lost since infall *and* the initial profile (encoded by $c_{\rm s}$), the evolution is indeed independent of the details of the stripping (which depends on the external potential, encoded by $c_{\rm h}$, and the subhalo's orbit, encoded by $x_{\rm c}$ and η).

Both H03 and P10 find that tidal evolution modifies the subhalo density profile in two main ways: (i) the outer density profile begins to drop off much more steeply with radius, transitioning from the $\mathrm{d}\log\rho/\mathrm{d}\log r = -3$ that is characteristic of the NFW profile at infall to $\mathrm{d}\log\rho/\mathrm{d}\log r = -(5-6)$, and (ii) the central densities slowly decrease with time as more and more mass is stripped away. The latter is mainly a consequence of the subhalo re-virializing in response to its mass loss. In addition, some of the reduction in central density arises more directly from the stripping of particles on highly eccentric orbits, which contribute mass to both the center and the outskirts. The impact of tidal shocking on the central densities is negligible as the short dynamical times in the dense centers imply adiabatic shielding (Gnedin & Ostriker, 1999; van den Bosch et al., 2018). Informed by these previous findings, and considering the newly-identified $c_{\rm s}$-dependence, we seek to describe the evolution of the subhalo density profile in terms of a transfer function $H(r|f_{\rm b}, c_{\rm s})$ that depends both on the initial subhalo concentration and the fraction of mass that has been stripped since infall.

Thus, the model-building procedure is largely one of exploratory data analysis and optimization. For calibrating candidate models of $H(r|f_{\rm b}, c_{\rm s})$, we employ a cost function that is the sum of squared logarith-

mic residuals between the DASH transfer functions and those predicted by the model:

$$E(\boldsymbol{\theta}) = \sum_{i}^{N_{\rm sim}} \sum_{j}^{N_{\rm snap}} \sum_{k}^{N_{\rm rad}} \Big\{ \log\left[H_{\rm D}(r_k|t_j, \{c_{\rm h}, c_{\rm s}, x_{\rm c}, \eta\}_i)\right] - \log\left[H_{\rm m}(r_k|f_{\rm b}(t_j), c_{\rm s}, \boldsymbol{\theta})\right]\Big\}^2 \tag{2.4}$$

Here, $\boldsymbol{\theta}$ denotes the free parameters of the model, and the sums run over all $N_{\rm sim}$ simulations, $N_{\rm snap}$ snapshots, and $N_{\rm rad}$ radial bins included in the preprocessed calibration dataset. $H_{\rm D}$ denotes the DASH transfer functions, which are labeled by the orbital parameters and halo concentrations at infall, snapshot number, and radial bin. $H_{\rm m}$ denotes the model transfer function, which only depends on the radial bin, bound fraction, initial subhalo concentration, and free model parameters. The adaptive Nelder-Mead downhill simplex method (Gao & Han, 2012) is used for model optimization due to its reliability and generalization to high-dimensional parameter spaces.

The DASH database does not contain a flat distribution of simulations across $c_{\rm h}$, $c_{\rm s}$, $x_{\rm c}$, and η, but rather consists of proportionally more simulations in the regions of parameter space that are more probable. Furthermore, the snapshots present in our calibration dataset do not contain a flat distribution in $f_{\rm b}$, as there are far fewer snapshots of subhaloes with low $f_{\rm b}$ than for the highest values. Thus, by using our flat cost function, which weights all radial bins and all snapshots equally, the calibrated model will perform best in the regions of parameter space that are most commonly found in cosmological simulations.

After testing a variety of functional forms for $H(r|f_{\rm b}, c_{\rm s}, \boldsymbol{\theta})$, we find that the transfer function is quite well described by

$$H(r|f_{\rm b}, c_{\rm s}, \boldsymbol{\theta}) = \frac{\rho(r,t)}{\rho(r,t=0)} = \frac{f_{\rm te}}{1 + \left(\tilde{r}\left[\frac{\tilde{r}_{\rm vir,s} - \tilde{r}_{\rm te}}{\tilde{r}_{\rm vir,s}\tilde{r}_{\rm te}}\right]\right)^{\delta}}, \tag{2.5}$$

which is a generalized form of the transfer function used in H03, which is given by $H_{\rm H03}(r|f_{\rm b}) = f_{\rm te}[1 + (\tilde{r}/\tilde{r}_{\rm s})^3]^{-1}$. Here, $\tilde{r} = r/r_{\rm s}$, such that all radii that appear in the transfer function are normalized to the initial NFW scale radius. The transfer function model contains three parameters:

$$f_{\rm te} = f_{\rm b}^{a_1\left(\frac{c_{\rm s}}{10}\right)^{a_2}} c_{\rm s}^{a_3(1-f_{\rm b})^{a_4}}, \tag{2.6}$$

$$\tilde{r}_{\rm te} = \tilde{r}_{\rm vir,s} f_{\rm b}^{b_1\left(\frac{c_{\rm s}}{10}\right)^{b_2}} c_{\rm s}^{b_3(1-f_{\rm b})^{b_4}} \exp\left[b_5\left(\frac{c_{\rm s}}{10}\right)^{b_6}(1-f_{\rm b})\right], \tag{2.7}$$

and

$$\delta = c_0 f_{\rm b}^{c_1\left(\frac{c_{\rm s}}{10}\right)^{c_2}} c_{\rm s}^{c_3(1-f_{\rm b})^{c_4}}. \tag{2.8}$$

These parametrizations were motivated based on power series expansions in $\log(f_{\rm b})$ and $\log(c_{\rm s})$ for the logarithms of $f_{\rm te}$, $\tilde{r}_{\rm te}$, and δ. Additional coupling between $f_{\rm b}$ and $c_{\rm s}$ was added and the functional forms were further adjusted through trial and error in order to maximally reduce the cost function in equation (2.4).

Clearly, $f_{\rm te}$ describes how the normalization of the inner density profile evolves. The other two parameters describe the steepening of the outer density profile. The tidal truncation radius $\tilde{r}_{\rm te}$ is related to the radius where the power-law begins to transition from NFW to a steeper, tidally stripped profile. The power-law slope at large radii is governed by δ, such that

$$H(r) \propto r^{-\delta} \implies \rho(r) \propto r^{-(3+\delta)} \quad \text{for} \quad \tilde{r} \gg \tilde{r}_{\rm te}. \tag{2.9}$$

This transfer function has several desirable, physically-motivated properties. Firstly, when $f_{\rm b} = 1$, the transfer function is unity for all radii, which is consistent with the fact that no tidal evolution has occurred yet. Furthermore, the truncation radius $r_{\rm te}$ starts at the virial radius and shrinks inwards only as the subhalo is tidally stripped.

Each of these three model parameters is itself parametrized to be a function of $c_{\rm s}$ and $f_{\rm b}$. In total, the 15 free parameters to calibrate are encoded in $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = \{a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, b_5, b_6, c_0, c_1, c_2, c_3, c_4\}. \tag{2.10}$$

We calibrate this model against the DASH simulations using the cost function and method described above, and the best fit parameters are listed in Table 2.1. Additionally, the dependence of the three functional parameters, $f_{\rm te}$, $r_{\rm te}$, and δ, on $f_{\rm b}$ and $c_{\rm s}$ can be seen in Fig. 2.2. Importantly, unlike the polynomial expansions used in H03, our power-law parametrizations of $f_{\rm te}$, $r_{\rm te}$, and δ are well-behaved down to arbitrarily low $f_{\rm b}$. Such a property will be crucial for using the model in a semi-analytic prescription for evolving subhalo populations, which, in the absence of an explicit mechanism for subhalo disruption, will continue to evolve subhaloes down to $f_{\rm b}$ below the resolution limit of DASH. For applications that do not depend on physically realistic extrapolation outside of the DASH $f_{\rm b}$ parameter space, an alternative, promising strategy for predicting the evolved subhalo density profile could involve employing a machine learning algorithm, such as random forest regression (Breiman, 2001). In agreement with previous works, our calibrated model demonstrates that the majority of the evolved subhalo density profiles are indeed well-described by $\rho \propto r^{-(5-6)}$ (i.e., $\delta \approx 2-3$). In particular, the outer density profile falls off more rapidly as subhalo concentration decreases.

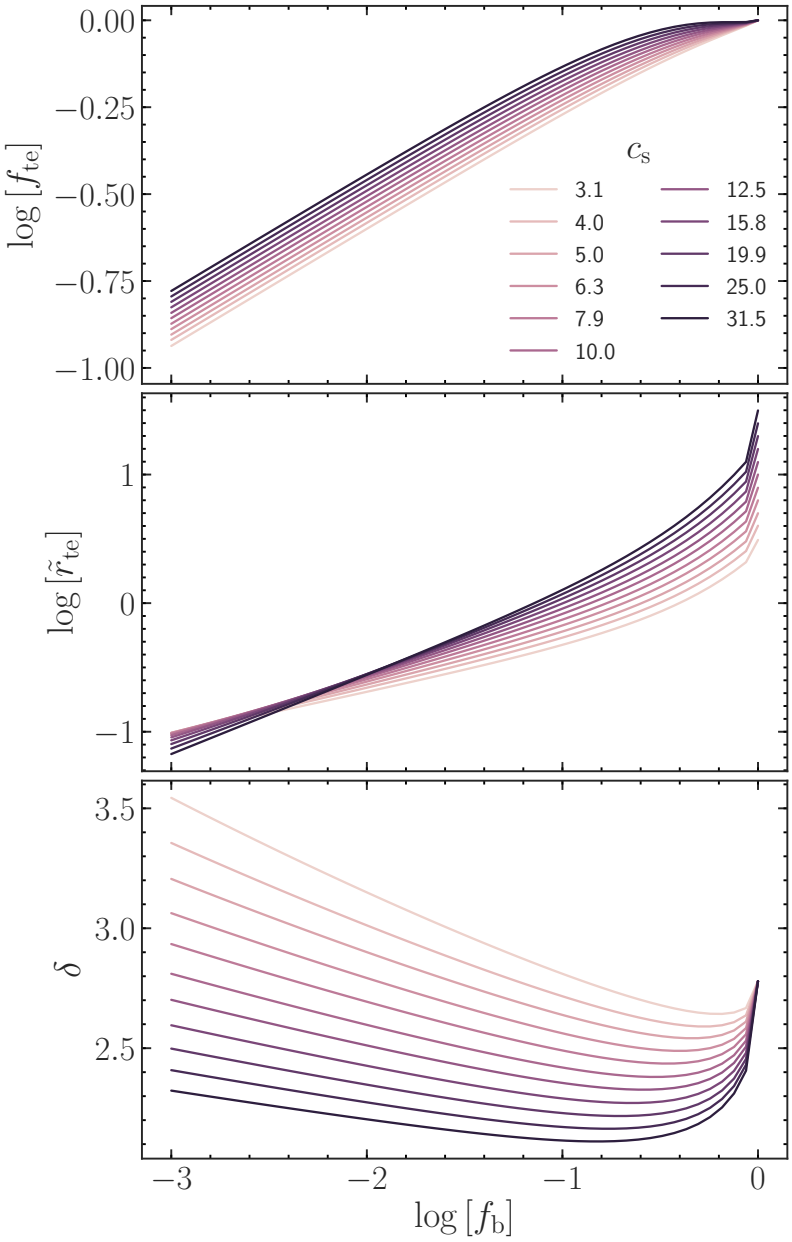

Figure 2.2: The dependence of the transfer function model (equation [2.5]) functional parameters (f_{te}, $\tilde{r}_{te}$, and δ, described by equations [2.6]–[2.8]) on the subhalo concentration c_s and bound fraction f_b. For the majority of the f_b-c_s parameter space, $\delta \approx 2 - 3$, resulting in a stripped subhalo density profile with $d\log\rho/d\log r = -(5-6)$, in agreement with previous idealized simulations (H03; P10). As the subhalo is increasingly stripped, δ increases and the outer profile drops off more steeply.

Since $\rho(r) \propto f_{te}$, the overall normalization of the density profile decreases as mass is stripped. The tidal truncation radius, r_{te}, roughly corresponds to the radius where the profile transitions to $d\log\rho/d\log r = -(3+\delta)$; this radius is smaller for subhaloes that are initially more concentrated.

a_1	0.338	b_1	0.448	c_0	2.779
a_2	0.000	b_2	0.272	c_1	−0.035
a_3	0.157	b_3	−0.199	c_2	−0.337
a_4	1.337	b_4	0.011	c_3	−0.099
		b_5	−1.119	c_4	0.415
		b_6	0.093		

Table 2.1: The best-fit parameters for the transfer function, $H(r|f_b, c_s)$ (equation [2.5]). These parameters are used to describe the dependence of the model's functional parameters (i.e., f_{te}, $\tilde{r}_{te}$, and δ, described by equations [2.6]–[2.8]) on the subhalo concentration c_s and bound fraction f_b. The best-fit value for parameter a_2 is consistent with zero, but the parameter was kept in order to maintain a consistent parametric form between f_{te} and the other functional parameters.

In Fig. 2.3, we compare our calibrated model to the DASH simulation transfer functions. Specifically, we first select a particular c_s, then bin the DASH simulation snapshots by f_b, which includes simulations over the parameter space of c_h, x_c, and η values. We plot these binned transfer functions versus radius, showing the medians and 16/84 percentiles for different ranges in f_b, as indicated. Our model transfer function is specified by c_s, f_b (which is equal to the f_b logarithmic bin center used for the DASH data), and the radius. The model demonstrates good agreement with the DASH simulation transfer functions across a large dynamic range in f_b and over the relevant c_s parameter space. To highlight our improved model and emphasize the benefits of using a large library such as DASH for data-driven model building, we overplot the transfer functions of H03 and P10. As described above, these models for the transfer function depend only on f_b. The model of P10, which was only calibrated to reproduce the structural parameters of subhaloes with $c_s = 23.1$, is able to capture the outer density profile of highly-stripped subhaloes with $c_s = 25$ quite well, whereas it fails to reproduce the corresponding inner density profiles. For the subhaloes with $c_s = 10$, the P10 model is better able to capture the inner density profile. The model of H03, which was calibrated only for subhaloes with $c_s = 10$, performs better for low c_s, but is not able to capture the inner profile normalization as well as our model, especially for highly-stripped haloes. An accurate model of the subhalo transfer function needs to depend on the initial density profile (encoded by c_s), as is clear from the fact that both the models of H03 and P10 perform much worse in the $c_s = 25$ case than in the $c_s = 10$ case. By incorporating dependence on c_s into our transfer function model, we are able to better reproduce the DASH simulation transfer functions for both example initial subhalo concentrations. We also emphasize the benefit of using a variable outer power law ($\delta \approx 2-3$) for the transfer function. In most cases, the outer slope of our transfer function model is bracketed by the values advocated in H03 ($\delta = 3$) and P10 ($\delta = 2$), enabling a more faithful reproduction of the outer profile across a broad range of f_b and c_s.

In Fig. 2.4, we plot the residuals between our model and the DASH simulation transfer functions, binned by radius and by each of f_b, c_s, c_h, x_c, and η. We find that there is no significant systematic correlation

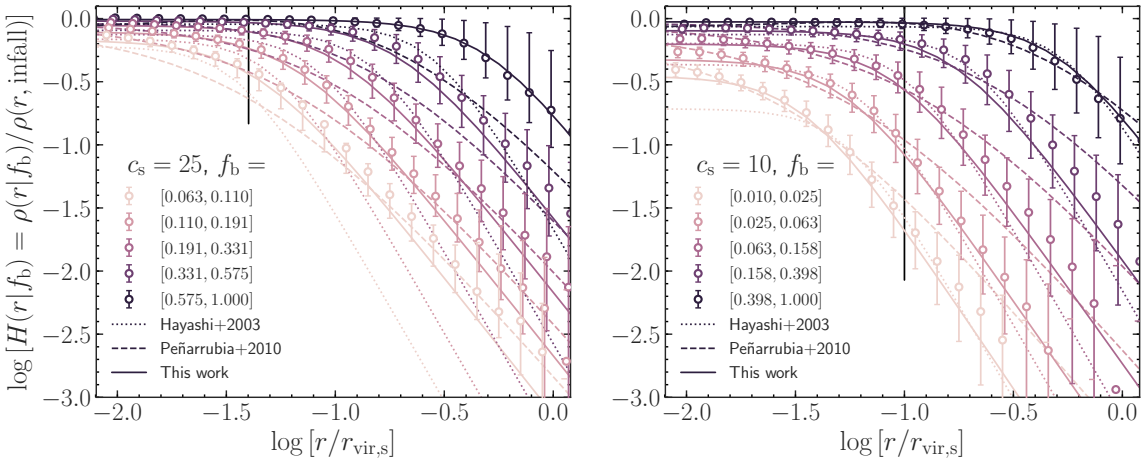

Figure 2.3: Comparison between evolved density profile models and the DASH simulations as a function of radius for $c_\mathrm{s} = 25$ (*left*) and $c_\mathrm{s} = 5$ (*right*), binned by f_b. All DASH simulation snapshots for subhaloes with the specified c_s that meet the preprocessing criteria (i.e., are numerically reliable and away from pericenter) and have f_b within the listed range are included. The open circles represent the median density profile transfer function value for the DASH simulations within the radial bin and f_b range, and error bars represent 16/84 percentiles. The f_b bins are progressively shifted horizontally for viewing in order to avoid overlapping error bars, but the true radii correspond to those of the lightest bin in f_b. The solid vertical lines denote r_s. For a large range of c_s values, the model accurately reproduces the tidally stripped subhalo density profile. The variable outer profile slope δ, described by equation (2.8), enables our model to better capture the outer density profile than H03 and P10, which use a fixed outer profile scaling of $\mathrm{d}\log\rho/\mathrm{d}\log r = -6$ or $\mathrm{d}\log\rho/\mathrm{d}\log r = -5$, respectively.

between the residuals and c_s or c_h. At the outer subhalo radii ($r \gtrsim 0.4 r_{\mathrm{vir,s}}$), the residuals increase for the most bound orbits (low x_c) and exhibit a weak dependence on η. Note also that the model is least accurate for the lowest bound mass fractions (i.e., $f_\mathrm{b} \lesssim 0.01$). Only a small fraction of all snapshots in DASH correspond to such small f_b values, all of which have small c_s. Consequently, this rare part of parameter space receives little weight in the optimization of the cost function, resulting in a less accurate fit. Note, though, that in each case the systematic offsets remain small compared to the halo-to-halo variance.

In the bottom right panel of Fig. 2.4, we give a final demonstration of the overall improvement of our model at reproducing the subhalo transfer functions of DASH compared to previous works. We plot the residuals between the various models and the DASH simulation transfer functions, now binned only by radius. These radial bins include all snapshots across the entire DASH dataset. Clearly, our updated prescription for the transfer function significantly improves upon previous work, as demonstrated by its nearly negligible bias at all radii and substantially reduced scatter. In particular, the addition of a variable power-law slope in the transfer function eliminates the strong bias at large radii seen in the residuals of the other two models. Thus, our model, calibrated on a massive dataset that is less prone to the numerical artifacts that plague cosmological simulations, provides the best predictions to date for the evolution of the subhalo

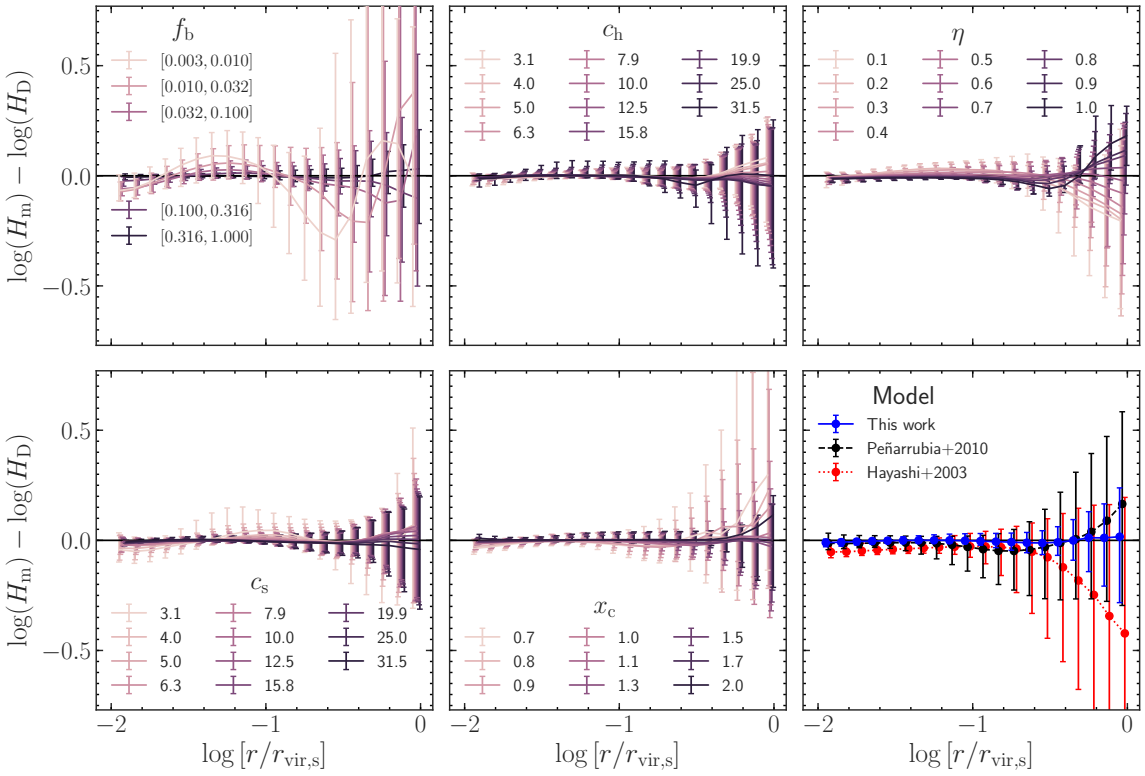

Figure 2.4: Residuals between the model and DASH simulation transfer functions (H_m and H_D, respectively) as a function of radius, binned by f_b (*top left*), c_s (*bottom left*), c_h (*top middle*), x_c (*bottom middle*), and η (*top right*). The *bottom right* plot compares the residuals between our model and the DASH simulations (blue dots) to the residuals between each of the models of H03 (red dots) and P10 (black dots) and the DASH simulations. Lines indicate the median residual and the error bars represent the 16/84 percentiles in each radial bin. The radii used for each value of the varied parameter (i.e., f_b, c_s, c_h, x_c, η, or the model) are progressively shifted horizontally for viewing, but the true radii correspond to those of the lightest-coloured curves (or the curve corresponding to 'this work' in the *bottom right* plot). See the text for a detailed discussion.

density profile. This tool will be a key ingredient in future semi-analytical models of dark matter substructure evolution.

2.4 Structural parameter evolution

Using the transfer function prescription developed above, one can easily compute the evolved subhalo density profile as $\rho(r|f_b, c_s) = H(r|f_b, c_s)\,\rho_{NFW}(r|c_s)$. Using the evolved profile, the radius of the maximum circular velocity, r_{max}, can be found by solving $r^3\rho(r) - \int_0^r r'^2\rho(r')dr' = 0$ for r. The associated maximum circular velocity is $V_{max} = \sqrt{GM(<r_{max})/r_{max}}$.

P10 find that the structural parameters of subhaloes, V_{max} and r_{max}, follow well-defined "tidal tracks" that only depend on f_b and the initial slope of the inner subhalo density profile. They calibrate a simple functional form for $V_{max}/V_{max,i}(f_b)$ and $r_{max}/r_{max,i}(f_b)$ based on their idealized subhalo simulations. They show that the functional form is accurate down to $f_b \approx 0.001$ in their simulations. While their simulations span a variety of initial inner density profile slopes, all simulated subhaloes have $c_s = 23.1$. On the other hand, H03, who only analyzed idealized subhalo simulations with $c_s = 10$, report that $V_{max} \propto f_b^{1/3}$; this result is inconsistent with the large cosmological simulations analyzed in Jiang & van den Bosch (2016) (see their Fig. 3) and, as we show below, is also inconsistent with DASH. The transfer function $H_{H03}(r|c_s)$ of H03 can also be used to calculate the evolution of the subhalo structural parameters, yielding a different relation that is more consistent with other models and the DASH data.

In addition to reproducing the evolved subhalo density profile, the performance of the model can also be quantified by its ability to reproduce the evolved structural parameters. For the initial values, we use the structural parameters of an NFW halo: $V_{max,i} = 0.465V_{vir}\sqrt{c/f(c)}$ and $r_{max,i} = 2.163r_s$ (here, $f(c) = \ln[1+c] - c/[1+c]$). In order to reduce the computational load of this analysis, we restrict ourselves to only the snapshots at apocentric passage, which still provides between 2–12 data points per simulation in the DASH database and a total of ~9,000 snapshots. For each snapshot, we compute the empirical structural parameters using the enclosed mass profile stored in DASH. The circular velocity profile is computed for each radial bin as $V_c(r) = \sqrt{GM(<r)/r}$ and then the structural parameters are determined from a fourth-order spline interpolation of this profile. Using each snapshot's associated values of f_b and c_s, the model predictions are calculated using the method described at the start of this section for our prescription and the one of H03. The predictions of P10 can be computed directly from their "tidal track" formula (their equation [8]).

In Fig. 2.5, we compare the model predictions for the structural parameters to the DASH results. Our model accuracy has minimal dependence on the stripped fraction, as evidenced by a similar level of scatter

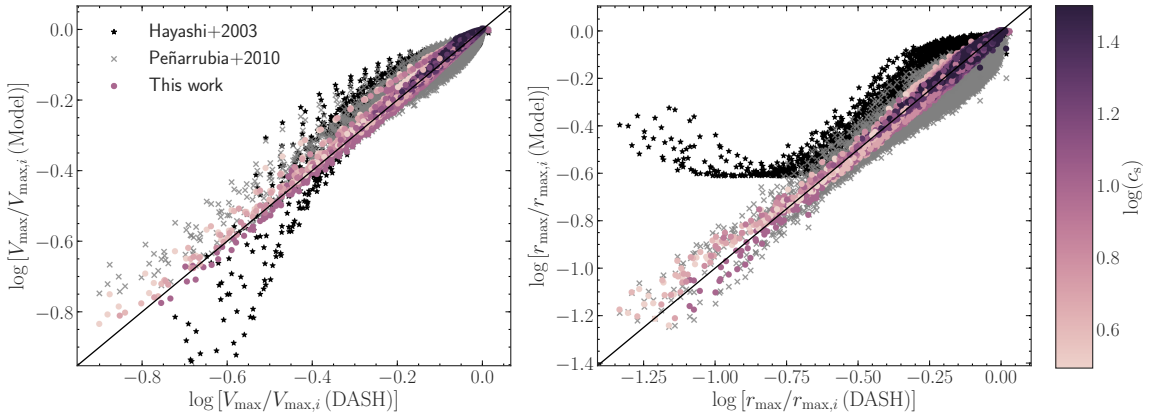

Figure 2.5: Scatter plots comparing the model predictions of the structural parameters normalized by their initial values, $V_{max}/V_{max,i}$ (*left*) and $r_{max}/r_{max,i}$ (*right*), to those of all DASH subhaloes at their apocentric passages. The results of our model are coloured by (the logarithm of) the subhalo concentration, demonstrating that the prediction's accuracy has minimal dependence on c_s. For comparison, the corresponding predictions from the models of H03 and P10 are also plotted (black stars and gray crosses, respectively), highlighting their increased scatter.

down to low $V_{max}/V_{max,i}$ and $r_{max}/r_{max,i}$. Additionally, the accuracy of our structural parameter predictions exhibits no residual dependence on the initial subhalo concentration. We overplot the predictions of H03 and P10, highlighting the significant improvement made by our model. In particular, much of the additional scatter in these prior models is due to the lack of c_s-dependence, which we illustrate below. In Fig. 2.6, we plot the DASH structural parameters against f_b, coloured by the initial subhalo concentration. This plot demonstrates that at fixed f_b, both V_{max} and r_{max} are larger for greater c_s, a trend that is exquisitely captured by our model due to the addition of c_s-dependence in the transfer function. A comparison between our model and the $H_{H03}(r|f_b)$-based structural parameter predictions illustrates the importance of using power law-based parametrizations in f_b. By parametrizing the model's functional parameters (f_{te}, r_{te}, and δ) as power laws in f_b and c_s, the transfer function and structural parameter predictions are well-behaved down to arbitrarily low f_b, unlike the model of H03, which uses a fitting function that is a polynomial expansion in $\log(f_b)$.

Overall, our model's ability to accurately reproduce the evolved subhalo density profiles and associated structural parameters across a wide range of subhalo parameter space represents an important step towards building a more accurate model of dark matter substructure evolution.

In order to aid the building of such models, we provide additional fitting functions for $V_{max}/V_{max,i}$ and $r_{max}/r_{max,i}$. We use the same "tidal track" formula introduced in Peñarrubia et al. (2008) and used in P10:

$$X(f_b, c_s) = \frac{2^\mu f_b^\eta}{(1+f_b)^\mu}, \tag{2.11}$$

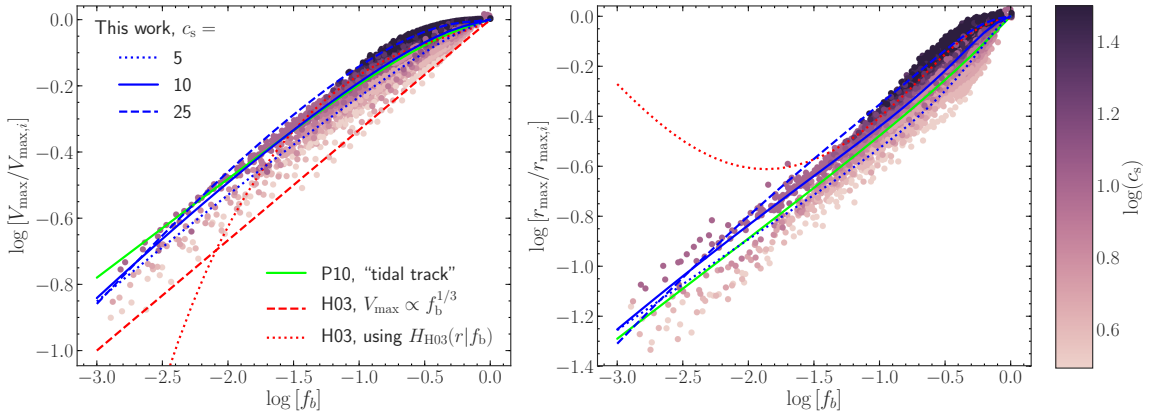

Figure 2.6: Scatter plots of $V_{max}/V_{max,i}$ (*left*) and $r_{max}/r_{max,i}$ (*right*) from all DASH subhalo snapshots at apocentric passages plotted against f_b and coloured by (the logarithm of) c_s. Overplotted are the model predictions of H03, P10, and this work (equations [2.12]-[2.13]). H03 reports that $V_{max} \propto f_b^{1/3}$ and also provides a transfer function $H_{H03}(r|c_s)$ that can be used to determine the structural parameters that results in a different relation. The latter are poorly behaved at small f_b due to the model's use of fitting functions that are polynomial expansions in $\log(f_b)$. The P10 predictions come directly from their "tidal track" fitting function (their equation [8]). The structural parameters can be determined using our transfer function model, which has dependence on c_s. As evidenced by the DASH data, such c_s-dependence is necessary in order to accurately capture the evolution of V_{max} and r_{max}.

where $\mu = \mu(f_b, c_s)$, $\eta = \eta(f_b, c_s)$, and X denotes either $V_{max}/V_{max,i}$ or $r_{max}/r_{max,i}$. P10 fit constants to each of the two functional parameters, μ and η; we introduce dependence on both f_b and c_s and instead write them as:

$$\mu(f_b, c_s) = p_0 + p_1 c_s^{p_2} \log(f_b) + p_3 c_s^{p_4}, \tag{2.12}$$

and

$$\eta(f_b, c_s) = q_0 + q_1 c_s^{q_2} \log(f_b). \tag{2.13}$$

The free parameters, $\boldsymbol{p}$ and $\boldsymbol{q}$, are fit to reproduce our model results for each of $V_{max}/V_{max,i}$ and $r_{max}/r_{max,i}$; the resulting values are listed in Table 2.2. The fitting function agrees with our model to $\lesssim$ 1% for $V_{max}/V_{max,i}$ and $\lesssim$ 3% for $r_{max}/r_{max,i}$ over the range $-3 \leq \log(f_b) \leq 0$ and $3.1 \leq c_s \leq 31.5$. Both the full transfer function model and the structural parameter fitting functions are well-behaved down to arbitrarily low f_b, which is a crucial characteristic for use in a semi-analytical model without disruption.

2.5 Summary and discussion

The evolution of dark matter haloes is predominantly studied through cosmological N-body simulations. These simulations show that haloes in virial equilibrium have universal density profiles (e.g., Navarro et al., 1997) and maintain a population of subhaloes that contain roughly 10% of the total halo mass (e.g., Ghigna

$V_{\rm max}/V_{{\rm max},i}$				$r_{\rm max}/r_{{\rm max},i}$			
p_0	2.980	q_0	0.176	p_0	1.021	q_0	−0.525
p_1	0.310	q_1	−0.008	p_1	1.463	q_1	−0.065
p_2	−0.223	q_2	0.452	p_2	0.099	q_2	0.083
p_3	−3.308			p_3	−4.643		
p_4	−0.079			p_4	−0.250		

Table 2.2: The parameters of the fitting function for the subhalo structural parameters normalized by their initial values, $V_{\rm max}/V_{{\rm max},i}$ and $r_{\rm max}/r_{{\rm max},i}$ (equation [2.11]), calibrated to agree with our transfer function model. These parameters encode the dependence of the model's functional parameters, μ and η, on the subhalo concentration $c_{\rm s}$ and bound fraction $f_{\rm b}$ (equations [2.12]–[2.13]).

et al., 1998; Gao et al., 2004; Giocoli et al., 2010). It has been shown that a large fraction of such subhaloes present in these simulations are completely disrupted within only a few orbital periods (Han et al., 2016; van den Bosch, 2017). Recently, several works have employed a combination of physical arguments and idealized simulations to claim that much of this subhalo disruption is artificial (Peñarrubia et al., 2010; van den Bosch et al., 2018; van den Bosch & Ogiya, 2018), indicating that the classical 'over-merging' problem (e.g., Katz & White, 1993; Moore et al., 1996) may still plague modern cosmological simulations. Specifically, van den Bosch & Ogiya (2018) showed that artificial disruption is primarily due to discreteness noise and inadequate force softening, a numerical issue that has been able to elude standard convergence tests. Hence, alternative approaches to studying the statistics of dark matter substructure are essential in order to cross-check the results of state-of-the-art simulations; only this will guarantee our ability to extract maximum information content that can be used for constraining the nature of dark matter and furthering the small-scale cosmology program.

As a promising alternative to N-body simulations, the semi-analytical modeling approach combines analytical halo merger trees, built using extended Press-Schechter theory (Bond et al., 1991), with a prescription for the tidal evolution of individual subhaloes as they orbit their host. This approach has been employed in a variety of previous models of substructure evolution (Taylor & Babul, 2001; van den Bosch et al., 2005b; Peñarrubia & Benson, 2005; Zentner et al., 2005a; Diemand et al., 2007; Kampakoglou & Benson, 2007; Gan et al., 2010; Pullen et al., 2014; Jiang & van den Bosch, 2016). These benefit from not being directly obstructed by the same numerical issues present in cosmological simulations. However, due to the lack of a fully analytical description of tidal evolution, these models still must be calibrated in some way against cosmological simulations (hence *semi*-analytical). The free parameters of the model are typically determined by tuning the results to reproduce the empirical subhalo mass functions of cosmological simulations. Clearly, if a large fraction of subhaloes in the simulations are subject to spurious disruption, then the semi-analytical models are calibrated against artificially suppressed subhalo mass functions, ultimately inheriting the same inadequacies of the simulations.

In an attempt to circumvent this issue, Ogiya et al. (2019) introduced the DASH subhalo evolution database, a suite of 2,253 idealized, high-resolution N-body simulations of individual subhaloes orbiting within a static, analytical host halo. These simulations are unimpaired by artificial disruption, with over 99.5% of the roughly 6×10^5 snapshots in the database passing the conservative numerical reliability criteria of van den Bosch & Ogiya (2018). The library samples the entire region of parameter space (i.e., initial orbital configurations and host-/subhalo concentrations) consistent with dark matter substructure observed in cosmological simulations.

This work represents the first phase of a research program devoted to building a semi-analytical model of dark matter substructure evolution that is calibrated against the DASH database and thus unobstructed by artificial disruption. In particular, this program will enable a calculation of the evolved subhalo mass function that is entirely independent of cosmological simulations, yielding a powerful method for validating the (small-scale) results of such simulations. In this chapter, we present an updated prescription for the evolution of the subhalo density profile. Previous such models by H03 and P10 only depend on the fraction of matter that has become unbound from the subhalo since infall (described by f_b). We find that the residuals between these f_b-only models and the DASH subhalo density profiles correlate significantly with the subhalo concentration c_s. Hence, we propose a more general model that depends both on f_b and the initial profile at infall (described by c_s). This evolved subhalo density profile is described by the transfer function $H(r|f_\mathrm{b}, c_\mathrm{s}) = \rho(r,t)/\rho(r,t=0)$, where we assume $\rho(r,t=0) = \rho_\mathrm{NFW}(r)$. Our model of this transfer function can be easily implemented in future semi-analytical models, as it has a simple algebraic form and is described fully by a set of parameters calibrated against the DASH simulations (see equations [2.5]-[2.8] and Table 2.1). As demonstrated in Section 2.3 and Section 2.4, our model is able to reproduce far more accurately the density profiles and structural parameters of evolved subhaloes than the models of previous work. In addition, we provide a fitting function for the evolving structural parameters, described by equations (2.11)–(2.13) and Table 2.2.

In the next chapter in this series (Section 3, which is published in Green et al., 2021a), we utilize the DASH library and our prescription for the evolved subhalo density profile to build a simple, physically-motivated model of the mass evolution of dark matter subhaloes. We will then combine this subhalo evolution model with accurate halo merger trees (e.g., Parkinson et al., 2008; Jiang & van den Bosch, 2014) to predict the evolved subhalo mass function of CDM haloes, a result that is completely free from the effects of artificial disruption. This will allow us to verify the predictions of Fig. 2.1 and determine whether or not the subhalo mass and velocity functions have indeed been severely underestimated. The results of this upcoming work will serve as an important check on the reliability of subhalo statistics derived from state-of-the-art cosmological simulations.

Chapter 3

Impact of artificial disruption

This chapter has been published as an article by Green et al. (2021a) in the Monthly Notices of the Royal Astronomical Society by Oxford University Press.

3.1 Background

THE standard Λ cold dark matter (ΛCDM) cosmological model predicts that structure forms as the consequence of primordial dark matter overdensities that collapse to form self-bound haloes. Smaller perturbations collapse earlier and merge to form larger haloes, resulting in a hierarchical halo assembly process that spans all mass scales. By studying halo evolution via cosmological N-body simulations, it is clear that the tightly bound central regions of smaller haloes survive the merger process, persisting as orbiting subhaloes within the treacherous environment of their host halo, where they are subjected to dynamical friction and disruptive tidal forces (e.g., Mo et al., 2010). Neglecting the impact of baryonic physics, this merger process is roughly self-similar due to the scale-free nature of gravitational collapse, ultimately resulting in an entire hierarchy of substructure, where subhaloes themselves host sub-subhaloes, and so on all the way down (Tormen et al., 1997; Gao et al., 2004; Kravtsov et al., 2004; Giocoli et al., 2010).

The population statistics of dark matter (DM) substructure are most often summarized in terms of subhalo mass functions (SHMFs) and radial profiles; these summary statistics depend heavily on the underlying particle nature of DM. For example, the free-streaming cutoff scale, set by the DM thermal velocity, impacts the low-mass end of the SHMF (e.g., Knebe et al., 2008; Lovell et al., 2014; Colín et al., 2015; Bose et al., 2017), while non-negligible DM self-interactions result in cored inner halo density profiles (e.g., Burkert, 2000; Vogelsberger et al., 2012; Rocha et al., 2013), which impacts the survivability of substructure in the presence

of tides (e.g., Peñarrubia et al., 2010). The predictions of substructure demographics made by these various dark matter models differ primarily at the low mass end. Consequently, many observational searches are underway in the attempt to constrain the abundance of low mass substructure, leveraging gravitational lensing (e.g., Dalal & Kochanek, 2002; Keeton & Moustakas, 2009; Vegetti et al., 2014; Shu et al., 2015; Hezaveh et al., 2016; Gilman et al., 2020b; Vattis et al., 2020), indirect detection via DM annihilation to γ-rays or decay signals (e.g., Strigari et al., 2007; Pieri et al., 2008; Hayashi et al., 2016; Hiroshima et al., 2018; Delos, 2019; Facchinetti et al., 2020; Rico, 2020; Somalwar et al., 2021), and gaps in stellar streams (e.g., Carlberg, 2012; Ngan & Carlberg, 2014; Erkal et al., 2016; Bonaca et al., 2020; Necib et al., 2020), among other approaches. Since satellite galaxies are inferred to live within subhaloes, with their respective properties related via the galaxy-halo connection, DM substructure statistics are intimately connected to satellite galaxy abundances (e.g., Vale & Ostriker, 2006; Hearin et al., 2013; Behroozi et al., 2013b; Newton et al., 2018; Nadler et al., 2019, 2021, 2020) and thus can be used to constrain cosmology through their impact on small-scale clustering statistics (e.g., Benson et al., 2001; Berlind et al., 2003; van den Bosch et al., 2005a; Lange et al., 2019; van den Bosch et al., 2019). Clearly, accurately modeling the evolution of DM subhalo populations is a prerequisite for their use as a cosmological probe and as a tool to study the particle nature of dark matter. Unfortunately, since the evolution of DM substructure is highly non-linear, modeling all but the most idealized circumstances has proven analytically intractable. Thus, to date, cosmological N-body simulations have been the most common avenue used for studying the demographics of DM substructure.

In recent years, cosmological simulations have successfully and repeatedly passed an important convergence test: as resolution is varied, the SHMFs remain in agreement above the 50–100 particle limit (e.g., Springel et al., 2008; Onions et al., 2012; Knebe et al., 2013; van den Bosch & Jiang, 2016; Griffen et al., 2016; Ludlow et al., 2019). While this is promising, the physical correctness of cosmological simulations is not guaranteed by the convergence of mass functions alone. Using the state-of-the-art *Bolshoi* simulation (Klypin et al., 2011), van den Bosch (2017) found that the evolved SHMF of surviving subhaloes looks identical to the SHMF of disintegrated subhaloes, noting that total subhalo disruption is prevalent. The inferred disruption rates from various studies are extremely high, with roughly 55-65% (90%) of subhaloes accreted at $z = 1$ (2) being disrupted by the present day (Han et al., 2016; van den Bosch, 2017; Jiang & van den Bosch, 2017). Hayashi et al. (2003) has shown that the total binding energy of a halo that is instantaneously stripped down to a sufficiently small radius (encompassing roughly 5–10% of the original mass) can be positive; hence, the authors suggested that such systems could disrupt spontaneously. Motivated by this analysis, subsequent works have incorporated physical disruption via tidal stripping and heating into their models or used such an argument as a justification for their results (Zentner & Bullock, 2003; Taylor & Babul, 2004; Klypin et al., 2015; Garrison-Kimmel et al., 2017). Recently, however, van den Bosch et al. (2018)

demonstrated that the boundedness of a subhalo remnant does not depend solely on the total binding energy, but rather on the radial distribution of the binding energies of the constituent particles. In fact, by using idealized simulations with sufficiently high resolution, van den Bosch et al. (2018) showed that it is possible for a self-bound remnant to survive even after 99.9% of the original mass has been stripped. More broadly, the study used analytical arguments to show that neither tidal heating nor tidal stripping alone are capable of causing complete physical disruption of cuspy CDM subhaloes (consistent with earlier work by Peñarrubia et al., 2010). As a follow-up, van den Bosch & Ogiya (2018) ran a suite of idealized numerical simulations, concluding that disruption of N-body subhaloes in cosmological simulations is largely numerical in nature and can be primarily attributed to (i) discreteness noise caused by insufficient particle resolution and (ii) inadequate softening of gravitational forces (see Mansfield & Avestruz 2020 for a recent analysis of the impact of the force softening scale on various halo properties). In agreement with these findings, van den Bosch (2017) assessed that approximately 80% of subhalo disruption in the *Bolshoi* simulation is most likely numerical in nature (see Section 3.2.5 below for details).

If the majority of subhalo disruption in cosmological simulations is indeed numerical, the implications for small-scale cosmology and astrophysics are profound. For example, a disruption-driven reduction in subhalo statistics would result in systematic biases in predictions from subhalo abundance matching (e.g., Conroy et al., 2006; Vale & Ostriker, 2006; Guo et al., 2010; Hearin et al., 2013; Chaves-Montero et al., 2016). Semi-analytical models of galaxy and dark matter substructure evolution (e.g., Taylor & Babul, 2001; Peñarrubia & Benson, 2005; Zentner et al., 2005a; Diemand et al., 2007; Kampakoglou & Benson, 2007; Gan et al., 2010; Pullen et al., 2014; Jiang & van den Bosch, 2016; Benson, 2020; Jiang et al., 2021; Yang et al., 2020) have historically been calibrated to reproduce the results of cosmological simulations and thus end up having inherited any systematic issues present in such simulations. As a specific example, Jiang & van den Bosch (2016) constructed a semi-analytical model that accurately matches the statistics of subhaloes in the *Bolshoi* simulation by simply tuning an orbit-averaged mass-loss rate and including an empirical model of subhalo disruption that, by construction, reproduces the disruption demographics in the simulation. As shown in Green & van den Bosch (2019, hereafter GB19), in the absence of such disruption, the normalization of the evolved SHMF predictions from Jiang & van den Bosch (2016) is boosted by a factor of two. Hence, depending on the fraction of subhalo disruption in cosmological simulations that is indeed artificial, it remains possible that such simulations (and derivative semi-analytical models) may be underestimating subhalo abundances by up to a factor of two. Such a systematic bias would have serious implications for dark matter indirect detection searches and could help explain the 'galaxy clustering crisis' in subhalo abundance matching (Campbell et al., 2018), since both of these applications, among others, depend heavily on evolved SHMFs from simulations. As long as the effects of artificial disruption remain as an unknown variable in the

analysis of cosmological simulations, we will be unable to extract the maximum amount of cosmological and astrophysical information content that will soon be made available in various large upcoming surveys, including DESI, LSST, EUCLID, and WFIRST. Clearly, there is still work to be done towards better understanding the tidal evolution of DM substructure, hence the motivation of the present study.

Recently, we released `SatGen` (Jiang et al., 2021), a semi-analytical modeling framework for studying galaxy and DM substructure evolution. The core components of the dark matter-only side of the framework include prescriptions for (i) analytical merger trees (Cole et al., 2000; Parkinson et al., 2008; Benson, 2017), from which the internal properties of subhaloes at accretion are derived, (ii) orbital parameter distributions for infalling subhaloes (Zentner et al., 2005a; Wetzel, 2011; Jiang et al., 2015; Li et al., 2020), (iii) the integration of subhalo orbits, including dynamical friction (Chandrasekhar, 1943), (iv) the evolved subhalo density profile, which captures how the internal structure of the subhalo responds to tidal heating and stripping (e.g., Hayashi et al., 2003; Peñarrubia et al., 2010; Drakos et al., 2017; Green & van den Bosch, 2019; Errani & Navarro, 2021), and (v) the instantaneous mass-loss rate, which depends on the structure of both the host- and subhalo in addition to the orbit (e.g., Drakos et al., 2020, this work). In contrast to Jiang & van den Bosch (2016), which followed van den Bosch et al. (2005b) by only considering orbit-averaged subhalo evolution, `SatGen` integrates individual subhalo orbits, thereby allowing for a proper treatment of splashback haloes (e.g., Ludlow et al., 2009; Aung et al., 2021; Diemer, 2021, 2020; Fong & Han, 2021). As we will show, this treatment of splashback haloes is crucial for properly comparing model predictions with simulation results.

The goal of this work is to build a semi-analytical model of substructure evolution that is independent of any tidal evolution-related numerical artifacts that may be present in cosmological simulations. Thus, in Ogiya et al. (2019), we introduced the Dynamical Aspects of SubHaloes (*DASH*) database, a large library of idealized, high-resolution N-body simulations of the tidal evolution of individual subhaloes. This simulation library has two key strengths: (i) the simulations span a wide range of parameter space, varying the initial orbital parameters and host- and subhalo concentrations and (ii) the live N-body subhaloes satisfy the strict set of convergence criteria laid out in van den Bosch & Ogiya (2018), suppressing numerical artifacts caused by discreteness noise and inadequate force softening. In GB19, we used *DASH* to calibrate a highly accurate, simply parametrized empirical model of the evolved subhalo density profile (ESHDP), which is unimpeded by numerical artifacts and is applicable to a far wider range of subhalo parameter space than that of previous works (Hayashi et al., 2003; Peñarrubia et al., 2010; Drakos et al., 2017). In this work, we use the results of GB19 as a component in a simple, physically motivated model of the instantaneous mass-loss rate. After calibrating this model to faithfully reproduce the subhalo mass trajectories across the range of *DASH* simulations, we incorporate it into `SatGen`, yielding the aforementioned artifact-free semi-analytical

model. We use this tool to make predictions for evolved subhalo mass functions, radial profiles, and substructure mass fractions and compare these findings to *Bolshoi* as an independent attempt to quantify the impact of artificial disruption on the abundance of dark matter subhaloes in cosmological simulations.

This chapter is organized as follows. In Section 3.2, we describe our methods, giving an overview of `SatGen` and our modifications, which include the incorporation of the Li et al. (2020) orbital parameter model (summarized in Section 3.2.2), the ESHDP model of GB19, and an improved, *DASH*-calibrated mass-loss rate. We also detail our procedures for modeling the impact of artificial disruption and calibrating the dynamical friction strength. In Section 3.3, we present the results of our augmented `SatGen` model, focusing on SHMFs, radial profiles, substructure mass fractions, and the numerical disruption rate in simulations. We conclude in Section 3.4 by summarizing our research program, highlighting the updates made to `SatGen`, and discussing our findings and their implications.

The cosmology used throughout this work is consistent with that of the *Bolshoi* simulation (Klypin et al., 2011): $\Omega_{\rm m} = 0.270$, $\Omega_\Lambda = 0.730$, $\Omega_{\rm b} = 0.0469$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_{\rm s} = 0.985$. The halo mass is defined as the mass enclosed within the virial radius, $r_{\rm vir}$, inside of which the mean density is equal to $\Delta_{\rm vir}(z)$ times the critical density. For the ΛCDM cosmology adopted in this work, $\Delta_{\rm vir}(z = 0) \approx 100$ and is otherwise well-described by the fitting formula presented by Bryan & Norman (1998). Throughout, we use m and M to denote subhalo and host halo masses, respectively. We use l and r to reference subhalo- and host halo-centric radii, respectively. Projected radii are indicated by upper-case letters. The base-10 logarithm is denoted by log and the natural logarithm is denoted by ln.

3.2 Methods

Our work builds on the original `SatGen` model that is presented in Jiang et al. (2021); we refer the reader to that paper for any additional model details that are omitted below. In what follows, we highlight the salient features of `SatGen` and discuss in greater detail the new modifications that we make as part of this study. Fig. 3.1 presents a schematic flowchart that summarizes all of the individual components of our framework.

3.2.1 Merger trees

Given an input that includes host halo virial mass, M_0, redshift of observation, z_0, and underlying cosmology, `SatGen` generates a user-defined number of halo merger trees that specify the subhalo masses and redshifts at which they are accreted by the main progenitor of each halo. Merger trees are constructed using the method of Parkinson et al. (2008), which is a modified version of the `GALFORM` 'binary method with accretion'

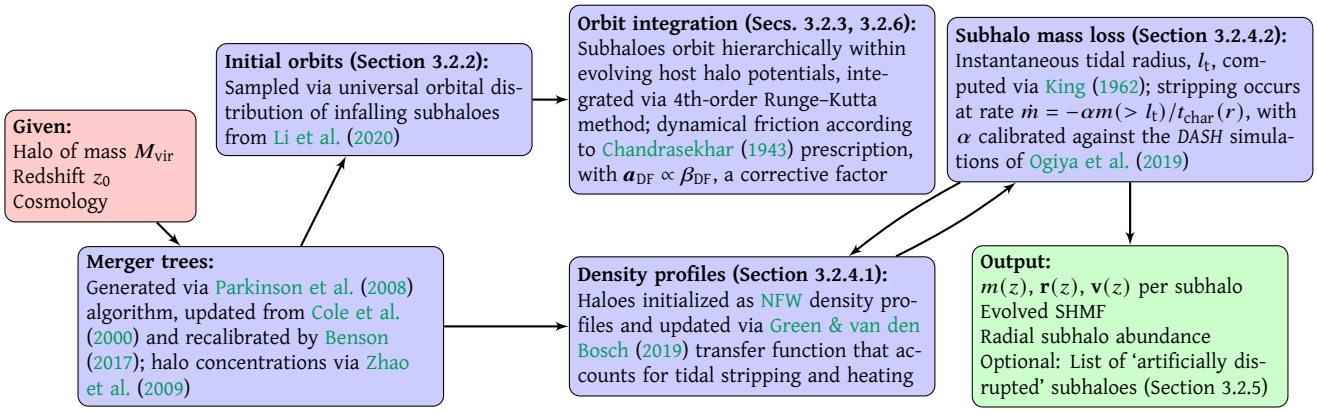

Figure 3.1: A flowchart that summarizes the `SatGen` framework employed in this study.

introduced by Cole et al. (2000). As demonstrated in Jiang & van den Bosch (2014) and van den Bosch et al. (2014), this method yields results that are in excellent agreement with numerical simulations.[1] As detailed in Jiang et al. (2021), we use the Parkinson et al. (2008) method with the updated set of parameters advocated for by Benson (2017) that are applicable to the Bryan & Norman (1998) virial halo mass definition. Each merger tree is characterized by a minimum progenitor mass, $M_{\rm res}$, which we set to be a fixed fraction, $\psi_{\rm res}$, of the final host halo mass, i.e., $M_{\rm res} = \psi_{\rm res} M_0$. The value of $\psi_{\rm res}$ used varies depending on the application and is specified accordingly. Following Parkinson et al. (2008), the merger tree is sampled using small time steps of $\Delta z \approx 10^{-3}$; however, in order to reduce memory usage, the tree is subsequently down-sampled to a temporal resolution of $\Delta t = \min[0.1 t_{\rm dyn}(z), 0.06\,{\rm Gyr}]$, where $t_{\rm dyn}(z) = \sqrt{3\pi/[16 G \Delta_{\rm vir}(z)\rho_{\rm c}(z)]}$ is the redshift-dependent halo dynamical time (see Jiang & van den Bosch, 2016). The maximum time step of 0.06 Gyr is motivated by convergence tests ran during the calibration of our subhalo mass-loss model, which we discuss in Section 3.2.4.2.

Both host haloes and subhaloes at accretion are assumed to follow a Navarro-Frenk-White density profile (hereafter NFW; Navarro et al., 1997) with a concentration parameter, $c_{\rm vir}$,[2] that depends on mass and redshift (or time) according to the model introduced by Zhao et al. (2009):

$$c_{\rm vir}(M_{\rm vir}, t) = 4.0 \left[1 + \left(\frac{t}{3.75 t_{0.04}} \right)^{8.4} \right]^{1/8} . \tag{3.1}$$

Thus, the concentration of the halo at a proper time, t, is determined based on the time at which its main

1. As an aside, we acknowledge that several components of the model, including the analytical merger tree algorithm and the orbital parameter distribution model, are still calibrated to agree with cosmological simulations. However, the calibration of these components only depend on properties of *unevolved* subhaloes (i.e., prior to accretion) and hence are not adversely impacted by any artifacts that may manifest in their subsequent tidal evolution.

2. We use $c_{\rm vir,h}$ and $c_{\rm vir,s}$ to refer to host- and subhalo concentrations, respectively.

progenitor has accumulated a mass of $0.04M_{\rm vir}(t)$, denoted $t_{0.04}$. Each branch of the merger tree has its own virial mass accretion history, $M_{\rm vir}(t)$, that tracks the halo from the time that it attains a mass of $0.04M_{\rm res} = 0.04\psi_{\rm res}M_0$ until the time that it merges into a more massive halo. Note that in order to have well-defined concentrations for all progenitor haloes down to a 'leaf mass' of $M_{\rm res}$, we track the main progenitor branch of each leaf further back in time down to $0.04M_{\rm res}$.

`SatGen` tracks subhaloes of *all orders*. The main branch, which follows the main progenitor of the $z = z_0$ host halo back in time, is considered to be order-0. Subhaloes that are directly accreted onto the main host are order-1. These subhaloes themselves can host sub-subhaloes, which are order-2, and so on. We use an inclusive mass definition in our merger trees, which means that the summed mass of all order-k subhaloes is included in the mass of their order-$(k-1)$ host. In some of our results (e.g., the SHMFs), we consider subhaloes of all orders; however, due to the inclusive mass definition, we only consider order-1 subhaloes for other results (e.g., the substructure mass fraction).

3.2.2 Initial orbits

Here, we describe our approach for initializing subhalo orbits. We specify the initial phase space coordinates of the infalling subhalo as

$$\{r, \theta, \phi, v_r, v_\theta, v_\phi\} = \{r_{\rm vir}, \theta, \phi, -v\cos\gamma, v\sin\gamma\cos\delta, v\sin\gamma\sin\delta\}. \tag{3.2}$$

We assume that subhalo infall occurs isotropically, and therefore select an initial azimuthal angle, ϕ, uniformly from $[0, 2\pi)$ and an initial polar angle, θ, by sampling $\cos\theta$ uniformly from $[0, 1)$. In order to determine the initial velocity vector, the degrees of freedom of which are the speed, v, the angle between the velocity vector and the (negative of the) position vector, γ, and an additional angle that sets the orientation of the orbital plane, δ, we use the universal model of Li et al. (2020), which has been calibrated using a large suite of cosmological simulations.

For all *first infall* events (i.e., for a given subhalo, only considering the first time a subhalo enters into the host virial radius) aggregated across all of the simulations and over a wide range of redshift snapshots, Li et al. (2020) find that $u \equiv v/V_{\rm vir,h}$ (here, $V_{\rm vir,h}$ denotes the instantaneous virial velocity of the host) is well-described by a universal log-normal distribution that is peaked near unity and is independent of subhalo mass and redshift, z:

$$p(u)\,{\rm d}u = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{\ln^2(u/\mu_1)}{2\sigma_1^2}\right]\frac{{\rm d}u}{u}. \tag{3.3}$$

Here, $\mu_1 = 1.2$ and $\sigma_1 = 0.2$. They also find that mergers with larger $M_{\rm vir}$ (instantaneous virial mass of

the host) and/or $\xi \equiv m_{\rm acc}/M_{\rm vir}$ result in more radial subhalo orbits, which is mainly attributed to gravitational focusing. By rewriting the host mass, $M_{\rm vir}$, in terms of its corresponding density peak height, $\nu \equiv \delta_{\rm c}(z)/\sigma(M_{\rm vir})$, where $\delta_{\rm c}(z)$ is the critical overdensity of collapse and $\sigma^2(M)$ is the mass variance, the authors find that the distribution of infall angles is redshift-independent and only depends on u, ν, and ξ. Specifically, $\cos^2\gamma$ follows an exponential distribution,

$$p(\cos^2\gamma \mid u, \nu, \xi)\, {\rm d}\cos^2\gamma = \frac{\zeta}{e^{\zeta}-1} \exp\left(\zeta \cos^2\gamma\right) {\rm d}\cos^2\gamma, \tag{3.4}$$

where

$$\begin{aligned} \zeta &= a_0 \exp\left[-\frac{\ln^2(u/\mu_2)}{2\sigma_1^2}\right] + A(u+1) + B, \\ A &= a_1\nu + a_2\zeta^c + a_3\nu\zeta^c, \text{ and} \\ B &= b_0 + b_1\zeta^c, \end{aligned} \tag{3.5}$$

and the best-fit parameters are $(a_0, a_1, a_2, a_3, b_0, b_1, \mu_2, c) = (0.89, 0.3, -3.33, 0.56, -1.44, 9.60, 1.04, 0.43)$.

We use equations (3.3)–(3.5) to sample the initial v and γ for each subhalo at infall. In order to set the orientation of the orbital plane, we assume isotropy and therefore draw δ uniformly from $[0, 2\pi)$.

Using the u distribution of Li et al. (2020) results in a substantial fraction of sampled orbits with initial orbital energies that lie above the maximum value sampled in the DASH simulations (corresponding to $x_{\rm c} = 2$). This fraction has a slight dependence on the host concentration. For example, for $c_{\rm vir,h} = 10$, a total of 25% of subhaloes have $x_{\rm c} > 2$ at infall and 2% are initially unbound (i.e., v at infall is larger than the escape velocity). Fortunately, the performance of our DASH-calibrated evolved subhalo density profile model and mass-loss prescription both exhibit minimal dependence on the orbital parameters. We emphasize that the combined impact of dynamical friction and the growth of the host potential results in continuous reduction of the subhalo orbital energy, lowering $x_{\rm c}$ over time. These effects also drive subhaloes that are initially unbound to eventually become bound after infall; thus, we include these initially unbound orbits in the `SatGen` subhalo population.

3.2.3 Orbit integration

Upon accretion, the initial orbital configuration (i.e., location on the virial sphere, orientation of the orbital plane, and the initial velocity vector) of each subhalo is drawn at random using the state-of-the-art universal infall model of Li et al. (2020, see Section 3.2.2 for details). Note that this is a significant and important improvement over the approach taken in the original `SatGen` paper, where it was assumed that all subhaloes

initially have an orbital energy of $E_{\rm orb} = V_{\rm vir,h}^2/2 + \Phi_{\rm h}(r_{\rm vir})$, where $V_{\rm vir,h}$ and $\Phi_{\rm h}$ are the instantaneous virial velocity and potential of the host halo, and a specific orbital angular momentum of $L_{\rm orb} = \eta r_{\rm vir} V_{\rm vir}$, where $\eta \in [0, 1]$ is drawn from a simple sinusoidal probability distribution, $p(\eta) = \pi \sin(\pi\eta)/2$.

Subhalo orbits are subsequently integrated according to the evolving potential of the immediate host and a simple prescription for dynamical friction. In particular, subhaloes are treated as point masses with phase space coordinates that are updated at each time step by integrating the following equation of motion:

$$\ddot{\boldsymbol{r}} = -\nabla\Phi_{\rm h} + \boldsymbol{a}_{\rm DF}. \tag{3.6}$$

The integration is performed using a fourth-order Runge–Kutta method. Here, $\boldsymbol{r}$ is the host-centric position vector of the subhalo and $\boldsymbol{a}_{\rm DF}$ is the acceleration due to dynamical friction (DF). The latter is modeled using the standard approach of Chandrasekhar (1943), which gives the acceleration as

$$\boldsymbol{a}_{\rm DF} = -4\pi G^2 m \ln\Lambda \rho(\boldsymbol{r}) F(<v) \frac{\boldsymbol{v}}{v^3} \tag{3.7}$$

(see Mo et al., 2010). Here, $\ln\Lambda = \ln(M/m)$ is the Coulomb logarithm, M and m are the instantaneous masses of the host and subhalo, respectively, $\rho(\boldsymbol{r})$ is the host NFW density profile, $\boldsymbol{v}$ is the relative velocity of the subhalo with respect to the host, and $F(<v)$ is the fraction of local host particles contributing to dynamical friction. The velocity distribution of the background particles is assumed to be Maxwellian and isotropic such that

$$F(<v_{\rm rel}) = \mathrm{erf}(X) - \frac{2X}{\sqrt{\pi}} \mathrm{e}^{-X^2}. \tag{3.8}$$

Here, $X \equiv v_{\rm rel}/(\sqrt{2}\sigma)$, where $\sigma(\boldsymbol{r})$ is the one-dimensional isotropic velocity dispersion of the host, which we compute using the Jeans equation for hydrostatic equilibrium in a spherical system (e.g., Binney & Tremaine, 2008). We use the orbital velocity of the subhalo for $\boldsymbol{v}_{\rm rel}$, ignoring the spin of the host halo.

Because of its simplicity and ability to produce results in reasonable agreement with simulations, equation (3.7) has long been the standard approach for capturing dynamical friction in semi-analytical models. However, it is based on a number of assumptions (i.e., a point particle moving in an isotropic, homogeneous background of field particles) that are clearly not justified when modeling the orbital evolution of dark matter subhaloes. In order to account for these (and other, see Mo et al., 2010) inherent shortcomings, we multiply $\boldsymbol{a}_{\rm DF}$ by a corrective factor, $\beta_{\rm DF}$, of order unity. We treat $\beta_{\rm DF}$ as a free parameter, which allows us to adjust the overall strength of dynamical friction (see Section 3.2.6).

3.2.4 Tidal stripping

As a subhalo orbits its host, it is subjected to tidal stripping and tidal shock heating. As discussed in detail in van den Bosch et al. (2018), neither of these processes can be rigorously treated analytically. Consequently, all previous semi-analytical models of subhalo evolution have calibrated their treatments using cosmological simulations, thereby inheriting any shortcomings present within such simulations (i.e., artificial disruption). The primary goal of this work is to build a semi-analytical model of DM substructure evolution that is calibrated in a way such that its results are not sensitive to such numerical artifacts. We achieve this by calibrating our model against *DASH*, a large suite of idealized, high-resolution N-body simulations that track individual, live N-body subhaloes as they orbit a fixed, analytical host halo potential (Ogiya et al., 2019). Both the host halo and the initial N-body subhalo in *DASH* are modeled as spherical NFW haloes. *DASH* consists of 2,253 simulations spanning a wide range of relevant parameter space, including initial orbital energy and angular momentum, as well as the concentration parameters of both the host- and subhalo. The library consists of various data products generated from each simulation, including the phase space coordinates of the subhalo centre-of-mass, the subhalo radial density profile, $\rho(l,t)$, and the bound mass fraction, $f_{\rm bound} \equiv m(t)/m_{\rm acc}$, where $m_{\rm acc}$ is the initial subhalo virial mass (i.e., the subhalo mass at accretion), each of which are recorded over 301 snapshots of time. Below, we use these results to calibrate a model that describes the evolution of the density profiles (Section 3.2.4.1) and bound masses (Section 3.2.4.2) of subhaloes as they orbit their host (note that the former is required for modeling the latter).

3.2.4.1 The evolved subhalo density profile (ESHDP)

In GB19, we used *DASH* to calibrate a model that describes how the internal structure of a subhalo evolves in response to tidal stripping and heating. In particular, motivated by the work of Hayashi et al. (2003) and Peñarrubia et al. (2010), GB19 present a 'transfer function' that describes the density profile of a tidally stripped subhalo as a function of its initial density profile and its instantaneous bound mass fraction, $f_{\rm bound}$. Consequently, the density profile of a subhalo at any time, t, is given by

$$\rho_{\rm s}(l,t) = H(l|\, f_{\rm bound}(t), c_{\rm vir,s})\, \rho_{\rm s}(l, t_{\rm acc}), \tag{3.9}$$

where $c_{\rm vir,s}$ is the concentration of the subhalo *at accretion* and $t_{\rm acc}$ denotes the time of accretion. The *DASH*-calibrated transfer function is given by

$$H(l|\, f_{\rm bound}, c_{\rm vir,s}) = \frac{f_{\rm te}}{1 + \left(\frac{l}{l_{\rm s}}\left[\frac{l_{\rm vir}-l_{\rm te}}{l_{\rm vir}l_{\rm te}}\right]\right)^{\delta}}. \tag{3.10}$$

Here, $f_{\rm te}$, $l_{\rm te}$, and δ are all expressed as fitting functions that depend on both $f_{\rm bound}$ and $c_{\rm vir,s}$ (see equations [6]–[8] and Table 1 of GB19; note that $l_{\rm te} \equiv r_{\rm te}$), whereas $l_{\rm s}$ and $l_{\rm vir}$ are the scale radius and virial radius of the NFW subhalo *at accretion*.[3] The transfer function describes how the outer density profile of the subhalo steepens from $\mathrm{d}\ln\rho/\mathrm{d}\ln l = -3$ (i.e., the outer slope of the initial NFW profile) to roughly $-(5-6)$ as the initial subhalo mass is stripped away. In addition, the central density of the subhalo is lowered as $f_{\rm bound}$ decreases, which is primarily a consequence of re-virialization in response to mass loss.

3.2.4.2 Mass-loss rate

A common approach to modeling the combined impact of tidal stripping and heating (e.g., Taffoni et al., 2003; Zentner & Bullock, 2003; Oguri & Lee, 2004; Zentner et al., 2005a; Pullen et al., 2014), which we adopt as well, is to assume that over each time step, Δt, some portion, Δm, of the subhalo mass outside of its instantaneous tidal radius, $l_{\rm t}$, is stripped away. In particular, we set

$$\Delta m = -\alpha \frac{\Delta t}{t_{\rm char}} m(>l_{\rm t})\,. \tag{3.11}$$

Here, α is a fudge factor that controls the stripping efficiency,

$$t_{\rm char} = \sqrt{\frac{3\pi}{16 G \bar{\rho}_{\rm h}(r)}} \tag{3.12}$$

is the characteristic orbital time of the subhalo (identical to the dynamical time introduced in Section 3.2.1), with r the instantaneous, host-centric radius of the subhalo and $\bar{\rho}_{\rm h}(r)$ the mean density of the host halo within r, and

$$l_{\rm t} = r\left[\frac{m(<l_{\rm t})/M(<r)}{2 + \frac{\Omega^2(t) r^3}{GM(<r)} - \frac{\mathrm{d}\ln M}{\mathrm{d}\ln r}\Big|_r}\right]^{1/3} \tag{3.13}$$

(King, 1962), with $\Omega(t) = |\mathbf{r}\times\boldsymbol{v}|/r^2$ the instantaneous angular orbital velocity of the subhalo. We have also experimented with other definitions of $t_{\rm char}$ and $l_{\rm t}$ but find that this combination, when used in conjunction

3. The dependence on $c_{\rm vir,s}$ went unnoticed in Hayashi et al. (2003) and Peñarrubia et al. (2010), both of which only studied subhaloes with a single concentration ($c_{\rm vir,s} = 10$ and 23.1, respectively).

with equation (3.11), is able to reproduce the *DASH* results most accurately.[4]

We use the $m(t)/m_{\rm acc}$ trajectories from the *DASH* simulations to calibrate α as follows. Given the data products from a particular *DASH* simulation, we create interpolators for $r(t)$, $\Omega(t)$, and $m(t)/m_{\rm acc}$. In order to avoid transient behavior in the simulations that results from the instantaneous introduction of a subhalo into its host potential (see Ogiya et al., 2019), we initialize our model based on the properties of the *DASH* subhalo at the beginning of its second orbit (i.e., after it has returned to apocentre for the first time). Given a choice for Δt and α, we evolve $m(t)/m_{\rm acc}$ using equation (3.11), where we set $m(> l_{\rm t}) = m(t) - m(< l_{\rm t})$. Here, $m(< l_{\rm t})$ is computed using the ESHDP of equation (3.9), which depends on the instantaneous value of $m(t)/m_{\rm acc}$ and the initial $c_{\rm vir,s}$,[5] and we demand that $\Delta m \geq 0$ such that the subhalo mass decreases monotonically. For each combination of simulation (indexed by i) and α value, we compute a cost function, $C(i|\alpha)$, which is simply the mean squared residual in $\log[m(t_j)/m_{\rm acc}]$ between our model and *DASH* averaged over all $n_{{\rm apo},i}$ apocentric passages subsequent to the initialization of our model (indexed by j). We then determine the total cost for a given α by computing the mean of the $C(i|\alpha)$ taken over all of the *DASH* simulations, which can be written explicitly as

$$C(\alpha) = \sum_{i}^{n_{\rm sim}} \frac{C(i|\alpha)}{n_{\rm sim}} = \sum_{i}^{n_{\rm sim}} \sum_{j}^{n_{{\rm apo},i}} \frac{\log^2[m_{{\rm model},i}(t_j)/m_{{\rm DASH},i}(t_j)]}{n_{\rm sim} n_{{\rm apo},i}}. \tag{3.14}$$

We emphasize that this cost function weighs each simulation equally, which is motivated by the fact that *DASH* samples the parameter space of orbits and halo concentrations according to a cosmological simulation-inferred joint probability distribution. The cost function depends somewhat on the time step used to integrate the model predictions (see equation [3.11]), but we find that the results converge with $\Delta t = 0.06$ Gyr, which we adopt throughout as the *maximum* time step for integrating the evolution of the subhalo in `SatGen`.

We find that $C(\alpha)$ is minimized for $\alpha \simeq 0.6$, for which the root-mean-square error in the apocentric mass predictions is 0.097 dex. In order to look for any secondary parametric dependence that the optimal α may have, we determine the best-fit α on a per-simulation basis, which we denote α_i. We then look at the correlation between α_i and the concentrations of the host- and subhalo as well as with the orbital parameters. We find that α_i depends strongly on $c_{\rm vir,s}/c_{\rm vir,h}$. By binning the simulations by $c_{\rm vir,s}/c_{\rm vir,h}$ and

4. The tidal radius is only an approximation of the zero-velocity surface, which itself is neither spherical nor infinitesimally thin, and different authors often adopt different definitions. See Read et al. (2006a), Tollet et al. (2017), and van den Bosch et al. (2018) for detailed discussions.

5. This enclosed mass profile is not analytical. Hence, in `SatGen`, we provide an interpolator for $m(< l)/m_{\rm acc}$ (and $\sigma(l)$, the one-dimensional isotropic velocity dispersion), which is itself a function of l, $m(t)/m_{\rm acc}$, and $c_{\rm vir,s}$. We interpolate over $\log[m(t)/m_{\rm acc}]$ and $\log(c_{\rm vir,s})$ using cubic B-splines and patch the surfaces together in $\log(l)$-space linearly.

taking the median α_i in each bin, we find a power-law relation that is well fit by

$$\alpha = 0.55\left(\frac{c_{\rm vir,s}/c_{\rm vir,h}}{2}\right)^{-1/3}. \tag{3.15}$$

This relation captures the fact that subhaloes that are more compact relative to their host are more resilient to stripping. We adopt this parametrization of α in `SatGen`, emphasizing that, for typical values of $c_{\rm vir,s}/c_{\rm vir,h}$, the concentration-dependence has a $\lesssim 30\%$ effect. In determining α, we use the instantaneous host $c_{\rm vir,h}$ (which evolves as long as the host itself has not yet become a subhalo) whereas the subhalo $c_{\rm vir,s}$ is fixed to its value at infall.

Although it is tempting to compare our best-fit value for α to that of previous semi-analytical models that rely on equation (3.11), such a comparison is frustrated by the fact that different studies have used different forms for $t_{\rm char}$ and/or $l_{\rm t}$ (see van den Bosch et al. 2018; Drakos et al. 2020 for detailed discussions). In addition, none of the previous studies have accounted for the detailed evolution of the subhalo density profile (as in, e.g., equation [3.9]), rendering such a comparison moot. We do emphasize, though, that by calibrating our model to the idealized *DASH* simulations, rather than to cosmological simulations, such as in Zentner et al. (2005a) and Pullen et al. (2014), our calibration is not adversely impacted by potential issues resulting from artificial disruption.

Fig. 3.2 compares the $m(t)/m_{\rm acc}$ trajectories of several *DASH* simulations (black lines) to predictions based on our mass-loss model (red lines). In each case, $c_{\rm vir,h} = 5$, $c_{\rm vir,s} = 10$, and the orbital energy, E, is that of a circular orbit at the virial radius of the host (i.e., $x_{\rm c} \equiv r_{\rm c}(E)/r_{\rm vir} = 1$, where $r_{\rm c}(E)$ is the radius of a circular orbit with energy E). Different panels correspond to different values of the orbital circularity, $\eta \equiv L/L_{\rm c}(E)$, as indicated, where L is the orbital angular momentum and $L_{\rm c}(E)$ is the angular momentum of a circular orbit with the same orbital energy as that of the subhalo. Clearly, our model tracks the *DASH* $m(t)/m_{\rm acc}$ curves quite faithfully over $\sim$5 radial orbital periods. Importantly, the performance of the model is strong over the full range of η, spanning from orbits that are close to radial ($\eta = 0.1$) to those that are close to circular ($\eta = 0.9$), a feat that has proven difficult for previous semi-analytical models of subhalo mass evolution (cf. Peñarrubia et al., 2010; Drakos et al., 2020). Although not shown, we emphasize that the model performs comparably for other configurations as well. In particular, the concentration dependence built into the parametrization of the stripping efficiency (i.e., equation [3.15]) considerably improves the predictions made for systems with $c_{\rm vir,s}/c_{\rm vir,h}$ ratios that deviate significantly from two.

We use the mass-loss model to predict the mass evolution of every simulated DASH subhalo. In Fig. 3.3, we plot the time evolution of the median and standard deviation of the log-residuals between our model predictions and the DASH mass trajectories. We find that the mass-loss model performs well over the full

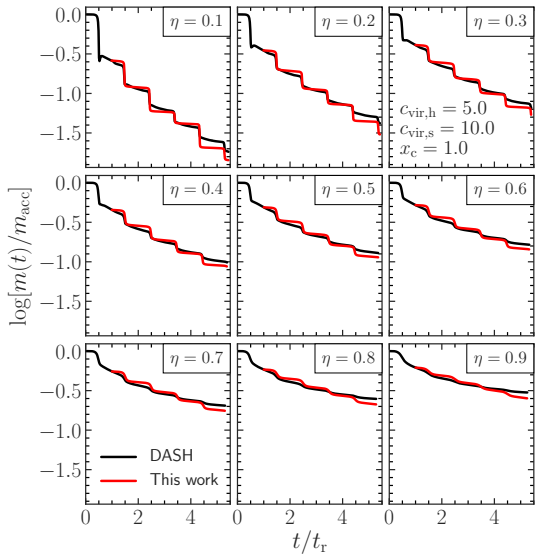

Figure 3.2: A comparison between our calibrated mass-loss model predictions and the *DASH* $m(t)/m_{\rm acc}$ trajectories of several simulations. The times are normalized by the radial orbital period, $t_{\rm r}$. We fix $c_{\rm vir,h} = 5$, $c_{\rm vir,s} = 10$, and $x_{\rm c} = 1$ (these parameters are typical of systems seen in cosmological simulations; see, e.g., Ogiya et al., 2019), demonstrating that the model performance is strong over a wide range of circularity values, ranging from highly elliptical ($\eta = 0.1$) to nearly circular orbits ($\eta = 0.9$).

parameter space, with minimal bias and scatter for longer than a Hubble time. After 15 Gyr of evolution, the scatter in the log-residuals of our mass-loss model reaches only 0.04 dex; hence, the impact of mass evolution error will be subdominant to the intrinsic halo-to-halo variance in our quantities of interest.

3.2.4.3 Stripping of higher-order substructure

In addition to the treatment of subhalo mass loss, `SatGen` also implements a procedure for the splashback release of higher-order subhaloes. Specifically, each time step that an order-k subhalo lies outside of the tidal radius of its order-$(k-1)$ host, it has a probability of $\min[\alpha \Delta t / t_{\rm char}(r), 1]$ of being released from its host and becoming an order-$(k-1)$ subhalo. Here, α and $t_{\rm char}(r)$ are computed for the order-$(k-1)$ host with respect to its order-$(k-2)$ parent, which is responsible for stripping off the order-k subhalo. In the event of release, the phase space coordinates of the subhalo with respect to its new, order-$(k-2)$ host are the superposition of its original coordinates with respect to its old, order-$(k-1)$ host and those of the old host with respect to the order-$(k-2)$ system. The remaining bound mass of the original, order-k subhalo is instantaneously removed from the mass of its old, order-$(k-1)$ parent in order to enforce mass conservation.

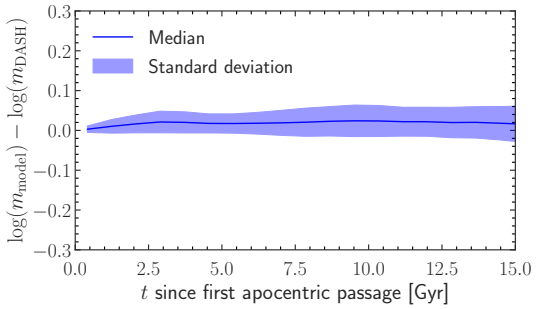

Figure 3.3: The time evolution of the median and standard deviation of the log-residuals between our mass-loss model predictions and the simulated mass trajectories taken over the ensemble of DASH simulations. The model performs well over the full parameter space, with minimal bias and scatter for longer than a Hubble time.

3.2.4.4 Resolution limits

As discussed in Section 3.2.1, `SatGen` has a merger tree resolution limit, which sets the smallest subhalo mass at accretion to $\psi_{\rm res} M_0$. Such a limit is necessary in order to maintain computational feasibility, as the size of the merger tree grows exponentially with decreasing $\psi_{\rm res}$. However, once accreted, a subhalo is evolved in `SatGen` for as long as its mass $m \geq \phi_{\rm res}\, m_{\rm acc}$. Here, $\phi_{\rm res}$ is the imposed resolution limit for the bound mass fraction. Our default is to set $\phi_{\rm res} = \psi_{\rm res}$, which ensures that the least (most) massive subhaloes are tracked down to $m = \phi_{\rm res}\psi_{\rm res} M_0$ ($m = \psi_{\rm res} M_0$). In what follows, both resolution limits are adjusted depending on the specific topic that is under investigation.

3.2.5 Artificial disruption

Recently, van den Bosch et al. (2018) and van den Bosch & Ogiya (2018) carried out a comprehensive analytical and numerical study focused on subhalo disruption. Using simple, physical arguments, the authors demonstrate that the inner remnant of a NFW subhalo should survive even when tidal shock heating has injected an amount of energy that is many multiples of the binding energy of the subhalo and/or tidal stripping has removed more than 99.9% of the initial subhalo mass. This claim is confirmed using idealized N-body simulations of subhalo evolution (similar to *DASH*), with the authors concluding that the majority of subhalo disruption seen in cosmological simulations is numerical in nature.

Let us use *Bolshoi* as our example cosmological simulation for considering the rate of artificial disruption. Van den Bosch (2017) used the merger trees from *Bolshoi* to separate subhalo evolution into several unique channels. Of these channels, the disruption (D) and withering (W) branches pertain specifically to numerical subhalo disruption. A subhalo in one snapshot that evolves along the D channel has no descendent at *any* subsequent snapshot. On the other hand, a subhalo that evolves along the W channel has a descendent in

the subsequent snapshot that falls below the 50 particle resolution limit imposed by the author. By studying these branches, van den Bosch (2017) concludes that artificial disruption (D) occurs at a rate of 2.4%/Gyr and falling below the mass limit (W) occurs at a rate of ~10%/Gyr. When combined, the total numerical disruption (W + D) rate in *Bolshoi* is roughly 13%/Gyr, resulting in ~65% of subhaloes accreted at $z = 1$ being numerically disrupted by the present day, in good agreement with independent estimates made by Han et al. (2016) and Jiang & van den Bosch (2017). As long as simulations have a finite number of particles, the W channel will exist. However, its significance diminishes as simulation resolution limits move toward smaller halo masses that are below all scales of interest. The D channel, on the other hand, is more alarming, since it represents subhaloes, often *well* above the mass limit, that simply disappear from the merger tree. The 2.4%/Gyr of the D channel translates to roughly 20% of subhaloes accreted at $z = 1$ being (artificially) disrupted by $z = 0$.

In order to assess the overall significance of numerical disruption, we aim to model both the impact of the W branch in isolation as well as the impact of both the W and D channels in combination on the `SatGen` results. As introduced in Section 3.2.4.2, the W branch subhaloes in `SatGen` are simply those with a final mass that has fallen below the merger tree resolution limit, $\psi_{\rm res} M_0$. Although even `SatGen` has an imposed resolution limit on how far down in $m/m_{\rm acc}$ it tracks a subhalo, we can nevertheless make reasonable predictions in the absence of withering by considering all subhaloes with $m/m_{\rm acc} \geq \phi_{\rm res} = 10^{-5}$, which we refer to as the "wither-free" fiducial model. Whenever withering is considered, the subhalo mass limit is set to $\psi_{\rm res} M_0$ instead.

A key goal of this work is to assess the impact of artificial disruption on the subhalo demographics in cosmological simulations. We are able to do so by adding a model of artificial disruption into `SatGen` and adjusting its strength (if needed) such that the `SatGen` predictions (which are inherently free of artificial disruption) reproduce the abundance of subhaloes in a simulation such as *Bolshoi*. This feat also requires properly accounting for the mass resolution limit (withering) of the simulation of interest. We implement a version of the artificial disruption mechanism used in Jiang & van den Bosch (2016), which itself is based on the prescription of Taylor & Babul (2004). A subhalo is marked as artificially disrupted when its mass, $m(t)$, falls below its 'disruption mass', given by

$$m_{\rm dis} = m_{\rm acc}(< f_{\rm dis} l_{\rm s}) = m_{\rm acc} \frac{f(f_{\rm dis})}{f(c_{\rm vir,s})}. \tag{3.16}$$

Here, $m_{\rm acc}(< l)$ denotes the enclosed NFW mass profile of the subhalo at accretion, and $f(x) = \ln(1+x) - x/(1+x)$. The sensitivity of haloes to artificial disruption is set by $f_{\rm dis}$, which represents the effective radius that a halo can be stripped down to before being disrupted. Under this prescription, haloes with a larger

initial concentration are more resilient to disruption. This approach to modeling (artificial) disruption has been employed in previous semi-analytical models (e.g., Hayashi et al., 2003; Taylor & Babul, 2004; Zentner et al., 2005a), with $f_{\rm dis}$ ranging from 0.1 to 2.0.

Rather than select a fixed value for $f_{\rm dis}$, Jiang & van den Bosch (2016) randomly sampled $f_{\rm dis}$ for each subhalo from a universal log-normal distribution. We augment this approach by calibrating a more general model of $f_{\rm dis}$ that takes into account a dependence on $m_{\rm acc}$ that we identify in the *Bolshoi* subhaloes. Using all halo catalogues from *Bolshoi*[6] with $z \geq 0.0148$,[7] we extract $m_{\rm acc}$, $c_{\rm vir,s}$, and $m_{\rm dis}$ from all D channel subhaloes from which the $f_{\rm dis}$ of each corresponding subhalo is calculated. We find that the distribution of $f_{\rm dis}$ has minimal dependence on redshift and host halo mass, but has a strong dependence on $m_{\rm acc}$. As shown in Fig. 3.4, when binned by $m_{\rm acc}$, the $f_{\rm dis}$ distribution is roughly log-normal with a log-mean, μ, and log-variance, σ^2, that increases and decreases, respectively, with decreasing $m_{\rm acc}$. This indicates that subhaloes that are more massive at accretion are less likely to undergo artificial disruption. However, note that this trend in $m_{\rm acc}$-space appears to saturate at the massive end. Motivated by these findings, we model $f_{\rm dis}(m_{\rm acc})$ as a log-normal with

$$\begin{aligned} \mu &= A + B\left[1 + \left(\log(m_{\rm acc}) + C\right)^{-2}\right]^{-1/2}, \quad \text{and} \\ \sigma &= D + E\mu + F\mu^2 . \end{aligned} \tag{3.17}$$

Using maximum likelihood estimation, we obtain the best-fit parameters of

$$(A, B, C, D, E, F) = (3.08, -3.26, -8.89, 0.38, -0.51, 0.40). \tag{3.18}$$

The corresponding best-fit model is indicated as solid lines in Fig. 3.4 and captures all of the salient details of the data.

When modeling artificial disruption in `SatGen`, we randomly draw a value of $f_{\rm dis}$ from the log-normal distribution described by equation (3.17) for each subhalo at accretion. Subsequently, the subhalo is marked as artificially disrupted once its mass drops below its assigned $m_{\rm dis}$, which is computed using equation (3.16). By applying this artificial disruption mechanism, `SatGen` is able to faithfully reproduce the statistics of the *Bolshoi* D branch subhaloes. We caution that this particular treatment of artificial disruption is only applicable to *Bolshoi*. Readers interested in modeling artificial disruption in another simulation must first characterize the corresponding $f_{\rm dis}$ statistics of the particular simulation.

6. Available at http://www.slac.stanford.edu/ behroozi/Bolshoi_Catalogs/

7. We omit using the several snapshots closer to $z = 0$ in order to avoid contaminating the D branch with instances of snapshot-limited failed phantom-patching (see discussion in van den Bosch, 2017).

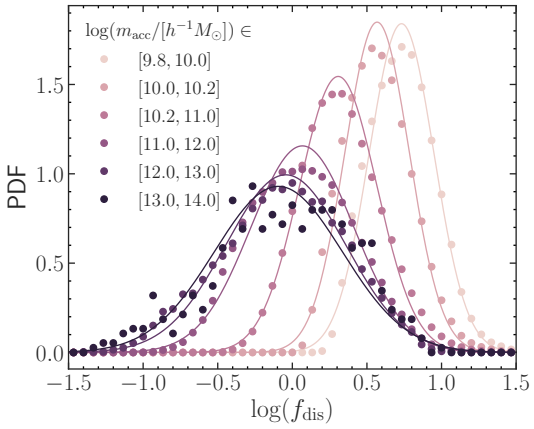

Figure 3.4: The $\log(f_{\rm dis})$ distribution of disrupted *Bolshoi* subhaloes. Here, $f_{\rm dis}$ is a proxy for the mass below which a particular subhalo is artificially disrupted in the simulation (see equation [3.16]). Each color denotes a different $m_{\rm acc}$ bin. The points are calculated using all *Bolshoi* subhaloes that disrupt at $z \geq 0.0148$. The solid curves correspond to our model that is fit to the *Bolshoi* disruption data (equation [3.17]). The $f_{\rm dis}$ are distributed log-normal, with μ decreasing (and σ increasing) as $m_{\rm acc}$ is increased (up to a saturation point, above which the distribution remains fixed).

3.2.6 Dynamical friction strength

In order to calibrate the overall efficiency of dynamical friction, which we quantify through the correction factor, $\beta_{\rm DF}$, we seek a measurement made from cosmological simulations that is both sensitive to dynamical friction and insensitive to any underlying artificial disruption. In van den Bosch et al. (2016), the authors study the segregation of subhaloes in *Bolshoi*. They measure the mean host-centric radius of subhaloes, $\langle r/r_{\rm vir}\rangle$, as a function of their redshift of accretion, $z_{\rm acc}$.

Plotting $\langle r/r_{\rm vir}\rangle$ (averaged over thousands of subhaloes) as a function of $z_{\rm acc}$ (see Fig. 7 in van den Bosch et al., 2016) reveals the characteristics of an orbit (for $z_{\rm acc} \lesssim 0.5$). Subhaloes accreted at $z_{\rm acc} \sim 0.1$ have just reached pericentre for the first time, while those that are at their first apocentric passage since accretion typically were accreted around $z_{\rm acc} \sim 0.25$. Note that phase mixing, which is primarily driven by variance in the orbital periods of subhaloes at infall,[8] results in a lack of orbital coherence for subhaloes accreted before $z_{\rm acc} \sim 0.5$; this is made apparent by the lack of clear apo- or pericentric passages in $\langle r/r_{\rm vir}\rangle$ at high $z_{\rm acc}$. Interestingly, the $\langle r/r_{\rm vir}\rangle(z_{\rm acc})$ curves show a clear dependence on $m_{\rm acc}/M_0$. In particular, subhaloes with larger $m_{\rm acc}/M_0$ reach a smaller apocentric $\langle r/r_{\rm vir}\rangle$ at $z_{\rm acc} \sim 0.25$ than their less massive counterparts (see Fig. 10 in van den Bosch et al., 2016, which is reproduced as the dashed lines in Fig. 3.5). This is a manifestation of dynamical friction, which allows us to calibrate $\beta_{\rm DF}$ as follows.[9]

8. The efficiency of phase mixing is further enhanced by dynamical friction, which impacts the subhalo orbit differently depending on $m_{\rm acc}$, and variance in the host mass accretion history, which itself affects the evolution of the subhalo orbit between infall and the present day.

9. Since artificial disruption is rare for subhaloes that were only accreted recently, this feature is not significantly impacted by artificial disruption.

We construct a set of ∼45,000 merger trees (with $\psi_{\rm res} = 10^{-3}$) with host masses consistent with the ∼9,000 host halo sample used in van den Bosch et al. (2016) — we augment our sample by generating five trees per unique host mass. We evolve the subhaloes with `SatGen`, repeating the procedure for several values of $\beta_{\rm DF}$ covering the range $[0, 1.5]$. We apply the same selection function as used in van den Bosch et al. (2016): we only consider subhaloes with $m(z=0)/M_0 \geq 10^{-3}$, $m_{\rm acc}/M_0 \geq 10^{-2}$, and $m(z=0)/m_{\rm acc} \geq 10^{-1}$. We first bin the subhaloes by $m_{\rm acc}/M_0$ and then compute $\langle r/r_{\rm vir}\rangle$ in $z_{\rm acc}$ bins, which are chosen such that the number of subhaloes in each bin is the same.

Fig. 3.5 shows the resulting $\langle r/r_{\rm vir}\rangle - z_{\rm acc}$ relation for four values of $\beta_{\rm DF}$ as indicated. Clearly, when $\beta_{\rm DF} = 0.75$, `SatGen` is able to very closely reproduce the simulation results. For $\beta_{\rm DF} = 0.5$ (1.0), `SatGen` yields apocentric $\langle r/r_{\rm vir}\rangle$ that are too large (small) relative to *Bolshoi*, with the disagreement being more significant for the subhaloes with larger $m_{\rm acc}/M_0$ that are more strongly influenced by dynamical friction. These findings are independent of whether or not we incorporate artificial disruption using the method described in Section 3.2.5, which is consistent with the notion that the $\langle r/r_{\rm vir}\rangle - z_{\rm acc}$ relation should be relatively insensitive to artificial disruption (at least for $z_{\rm acc} \lesssim 0.5$). Hence, in what follows, we adopt $\beta_{\rm DF} = 0.75$ as our fiducial dynamical friction strength. In Section 3.3.3, we quantify the impact of $\beta_{\rm DF}$ on our substructure mass fraction predictions (by comparing to the 'natural' case of $\beta_{\rm DF} = 1$), demonstrating that our results are insensitive to its exact value.

3.3 Results

Given a host halo mass, M_0, target redshift, z_0, and requested number of individual trees, $N_{\rm tree}$, `SatGen` produces $N_{\rm tree}$ subhalo catalogs at each redshift time step until z_0. These catalogs trace the mass and phase space coordinates of each subhalo over its evolution. In this section, we present the results of these `SatGen` subhalo catalogs and make comparisons to *Bolshoi*. We begin by studying SHMFs (and subhalo maximum circular velocity functions), comparing `SatGen` results with and without the artificial disruption mechanism and discuss the significant impact of splashback subhaloes (Section 3.3.1). In Section 3.3.2, we proceed to incorporate position data by calculating the radial profile and the (projected) enclosed substructure mass fraction, $F_{\rm sub}(<R)$. In Section 3.3.3, we quantify how $f_{\rm sub}(<r_{\rm vir})$ varies with both M_0 and resolution limit, $\psi_{\rm res}$. We also quantify the impact of model parameters (i.e., stripping efficiency and DF strength) on $f_{\rm sub}$ predictions. Lastly, in Section 3.3.4, we estimate the total rate of numerical disruption that occurs via the W and D channels modeled by `SatGen`, which we compare to the numerical disruption rate of *Bolshoi* haloes (as measured by van den Bosch, 2017).

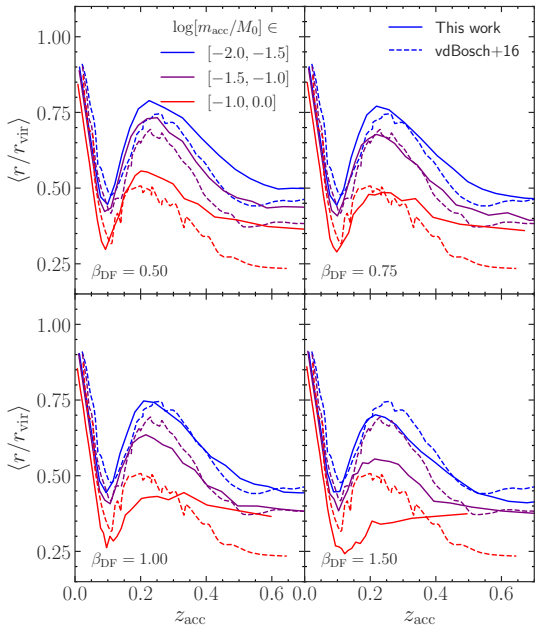

Figure 3.5: A comparison between the $\langle r/r_{\rm vir}\rangle - z_{\rm acc}$ relation of *Bolshoi* subhaloes binned by infall mass relative to $z = 0$ host mass, $m_{\rm acc}/M_0$ (dashed curves, reproduced from Fig. 10 in van den Bosch et al., 2016), and analogous predictions by `SatGen` (solid curves). Each panel corresponds to a different value of $\beta_{\rm DF}$ (indicated in the bottom-left of each panel), which controls the strength of dynamical friction (see Section 3.2.3). We adopt $\beta_{\rm DF} = 0.75$ as the fiducial value used in `SatGen`, since this yields the best agreement with respect to the peak values of $\langle r/r_{\rm vir}\rangle$ at $z_{\rm acc} \sim 0.25$, which corresponds to the first apocentric passage since infall.

3.3.1 Subhalo mass/velocity functions

We turn our attention to the `SatGen` predictions of the subhalo mass function for a $10^{14.2}\ h^{-1}{\rm M}_\odot$ host. In a cosmological simulation, the SHMF, $dN/d\log(m/M_0)$, is calculated using subhaloes *of all orders* enclosed within the virial radius of the host. Note that since we use an inclusive mass definition and consider all orders of substructure, the total substructure mass is *not* the mass-weighted integral of the SHMF. The left-hand panel of Fig. 3.6 shows the mean SHMF computed from 10,000 trees (with $\psi_{\rm res} = 10^{-4}$). For comparison, the filled symbols indicate the mean SHMF of the 282 *Bolshoi* host haloes with $\log(M_0/[\ h^{-1}{\rm M}_\odot]) \in [14.0, 14.5]$ (with a mean of 14.2). On the high-m/M_0 end, the *Bolshoi* SHMF is somewhat noisy due to limited halo statistics. However, a comparison at the low-m/M_0 end illustrates that `SatGen` predicts a ~ 0.1 dex enhancement in the SHMF relative to *Bolshoi*.

If the primary cause of the disagreement between the `SatGen` and *Bolshoi* SHMFs is artificial disruption, then the application of our artificial disruption mechanism (Section 3.2.5) should result in better agreement between the model and simulation results. Indeed, Fig. 3.6 shows that "turning on" D channel disruption suppresses the `SatGen` SHMF by $\sim 0.05-0.1$ dex at low m/M_0, bringing it into closer agreement with *Bolshoi*. Restricting to $-3.9 \leq \log(m/M_0) \leq -0.8$, we fit a power-law to the SHMF of the form $dN/d\log(m/M_0) \propto$

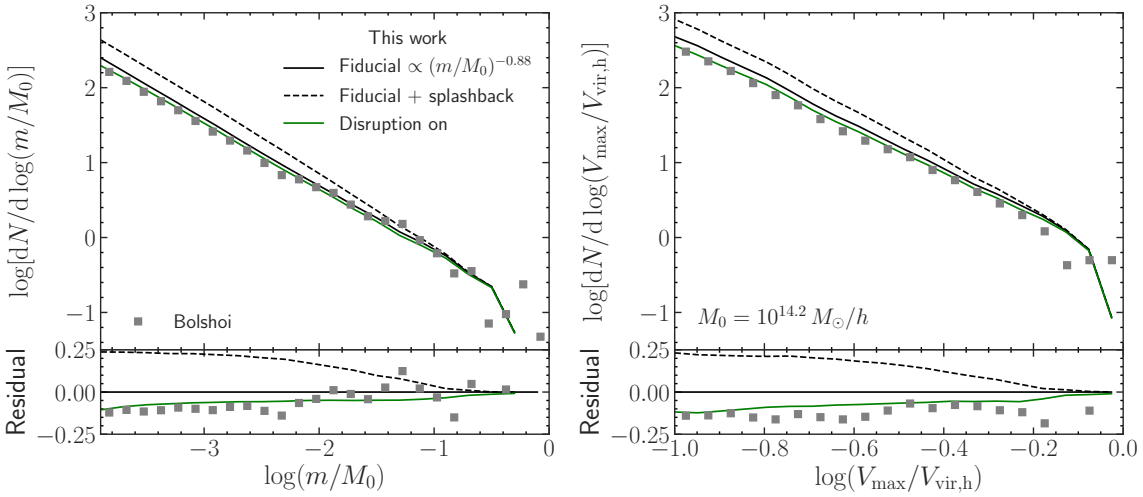

Figure 3.6: The subhalo mass function (SHMF; *left*) and subhalo maximum circular velocity function (SHVF; *right*) predictions for a host halo with $M_0 = 10^{14.2}\ h^{-1}\mathrm{M}_\odot$ at $z = 0$ (and virial velocity at $z = 0$ denoted by $V_{\rm vir,h}$). The `SatGen` results are averages taken over 10,000 merger trees generated with $\psi_{\rm res} = 10^{-4}$. These results are compared to the same quantities computed from 282 *Bolshoi* host haloes with $\log(M_0/[\ h^{-1}\mathrm{M}_\odot]) \in [14.0, 14.5]$ (with a mean of 14.2), which are shown as gray squares. The fiducial `SatGen` predictions (black lines) are used as the baseline for comparison in the residual plots. In the "fiducial + splashback" case (dashed black lines), we include subhaloes in the SHMF that are in the merger tree but instantaneously lie outside of the host $r_{\rm vir}$ at $z = 0$. Lastly, the "disruption on" case (green lines) demonstrates the impact of our artificial disruption mechanism (Section 3.2.5), which is calibrated to reproduce the statistical properties of *Bolshoi* subhalo disruption (D channel). At the low-m/M_0 end, artificial disruption suppresses the `SatGen` SHMF by $\sim 0.05 - 0.1$ dex, which brings our predictions into good agreement with *Bolshoi*. For $m/M_0 \lesssim 10^{-2.5}$, nearly half of the subhaloes lie outside $r_{\rm vir}$ (consistent with Bakels et al., 2021).

$A(m/M_0)^B$. For the fiducial SatGen results, we find $A = -1.066$ and $B = -0.885$, whereas the disruption mechanism slightly suppresses both the normalization and the magnitude of the slope, resulting in $A = -1.085$ and $B = -0.868$.[10] The reduced slope is a consequence of the $m_{\rm acc}$-dependence of our artificial disruption model. Note that the *Bolshoi* SHMF is too noisy to compute a reliable estimate of the slope over the same mass range, but it agrees well visually with the SatGen "disruption on" results.

Thus far, these results suggest that artificial disruption has, at most, a $\sim$ 20% impact on the SHMF of well-resolved host haloes, with the difference being strongest at low m/M_0. We discuss more quantitatively the impact of disruption and its dependence on halo mass relative to the simulation resolution limit in Section 3.3.3, which focuses on the substructure mass fraction. The modest decrease in the SHMF normalization due to disruption predicted by SatGen is considerably smaller than the factor-of-two suppression suggested by the GB19-interpretation of the Jiang & van den Bosch (2016) model. This is because their orbit-averaged model did not take into account the impact of splashback haloes, which are subhaloes that have previously fallen within the host $r_{\rm vir}$ (thus becoming included in the halo merger tree) but instantaneously lie outside of $r_{\rm vir}$ at $z = 0$ (and therefore are typically not included in simulation-based measurements of the SHMF). Benson (2017) briefly discusses this limitation of standard EPS-based approaches to substructure modeling, concluding that a full dynamical model (such as SatGen) is necessary in order to properly account for splashback haloes. In Fig. 3.6, we illustrate that when splashback haloes are included in the SHMF, the subhalo abundance is enhanced by $\sim 0.2 - 0.25$ dex on the low-m/M_0 end relative to the fiducial model. When the "fiducial + splashback" curve is compared directly to *Bolshoi*, we find the same ~ 0.3 dex (factor of two) difference as GB19. This highlights the importance of properly accounting for splashback haloes by integrating subhalo orbits. Consistent with these predictions, Bakels et al. (2021) recently reported that roughly half of all subhaloes lie outside of $1.2r_{\rm 200c}$ (approximately $r_{\rm vir}$) in a sample of galaxy- to group-mass host haloes studied in a cosmological simulation (consistent with previous work by, e.g., Gill et al., 2004; Ludlow et al., 2009).

In addition to mass, another property of subhaloes that is often used (especially in subhalo abundance matching, e.g., Trujillo-Gomez et al., 2011; Reddick et al., 2013; Hearin et al., 2013; Zentner et al., 2014) is its maximum circular velocity, $V_{\rm max}$. Hence, we also present results for the subhalo maximum circular velocity function (SHVF), $\mathrm{d}N/\mathrm{d}\ln(V_{\rm max}/V_{\rm vir,h})$, where $V_{\rm vir,h}$ denotes the virial velocity of the host halo at $z = 0$. Since the enclosed mass profile corresponding to the GB19 ESHDP is not analytical, we compute $V_{\rm max}$ by

10. These SHMF slopes are consistent with previous work, which typically find $-1.0 \lesssim B \lesssim -0.8$ (e.g., Boylan-Kolchin et al., 2010; Gao et al., 2012; van den Bosch & Jiang, 2016).

multiplying the subhalo's maximum circular velocity at accretion,

$$V_{\rm max,acc} = \sqrt{\frac{G m_{\rm acc}}{l_{\rm vir}} \times \frac{0.216\, c_{\rm vir,s}}{f(c_{\rm vir,s})}} \tag{3.19}$$

(Bullock et al., 2001), by the 'tidal track' (Peñarrubia et al., 2008) expression for $V_{\rm max}/V_{\rm max,acc}$, given by equation (11) in GB19. This tidal track itself is a function of both $m/m_{\rm acc}$ and $c_{\rm vir,s}$, as given by equations (12) and (13) in GB19.

Using the same 10,000 trees as those used to compute the SHMF, we obtain the SHVF shown in the right-hand panel of Fig. 3.6. Just as for the SHMF, the filled symbols indicate the corresponding result computed from the 282 *Bolshoi* host haloes with $\log(M_0/[\ h^{-1}{\rm M}_\odot]) \in [14.0, 14.5]$. As is evident, the abundance of subhaloes with $\log(V_{\rm max}/V_{\rm vir,h}) \lesssim -0.4$ predicted by `SatGen` is about 0.15 dex higher than that of *Bolshoi*. However, when including artificial disruption, the `SatGen` predictions once again agree closely with the simulation results.

3.3.2 Radial profiles

Having looked at the subhalo mass and velocity functions, we proceed to incorporate additional spatial information by considering several other quantities of interest. First, we measure the subhalo radial distribution, $d\tilde{N}/dx^3|_{\rm sub}$, as the number of subhaloes per unit shell volume as a function of $x = r/r_{\rm vir}$, which we normalize to unity at $r_{\rm vir}$. We assess the radial bias of the subhaloes by comparing $d\tilde{N}/dx^3|_{\rm sub}$ to the NFW profile of the host halo, $d\tilde{N}/dx^3|_{\rm NFW}$, which we also write as a function of x and normalize to unity at $r_{\rm vir}$. The 'bias function' is simply the ratio between $d\tilde{N}/dx^3|_{\rm sub}$ and $d\tilde{N}/dx^3|_{\rm NFW}$, which tends to unity when the subhalo distribution is unbiased with respect to the density profile of the host. We incorporate subhaloes of all orders when computing $d\tilde{N}/dx^3|_{\rm sub}$. The second quantity of interest is the fraction of mass enclosed within a given *projected* host-centric radius that is bound in subhaloes. We define this quantity as

$$F_{\rm sub}(<X) = \frac{1}{M(<X)} \sum_{X_i<X} m_i\,, \tag{3.20}$$

where $X = R/R_{\rm vir}$, R is the projected radius, the sum runs over all *first-order* subhaloes (due to the inclusive mass definition) with projected radii within R, and $M(<R)$ is the projected mass profile of the NFW host halo (see Golse & Kneib, 2002). Finally, in Section 3.3.3, we focus on $f_{\rm sub}(<r_{\rm vir})$, which is computed in the same way as $F_{\rm sub}$ except that three-dimensional radii are used instead.

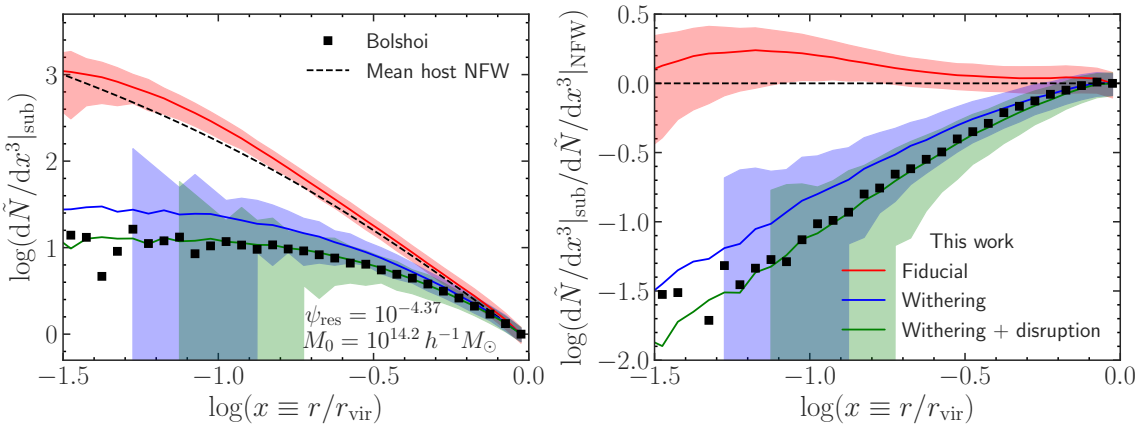

Figure 3.7: The subhalo radial profile (including subhaloes of all orders), $\mathrm{d}\tilde{N}/\mathrm{d}x^3|_{\rm sub}$ (*left*), normalized to unity at $r_{\rm vir}$ and the bias function (*right*), which quantifies how the radial profile differs from the host density profile. The `SatGen` results are computed from 2,000 merger trees of systems with $M_0 = 10^{14.2}\ h^{-1}\mathrm{M}_\odot$ at $z = 0$ and a merger tree resolution limit of $\psi_{\rm res} = 10^{-4.37}$, consistent with the *Bolshoi* resolution limit for hosts of the same mass. The lines represent the sample means and the shaded regions denote the 16–84 percentiles taken over the sample, which quantify the halo-to-halo variance. The fiducial result (red) includes all subhaloes with $m/m_{\rm acc} \geq 10^{-5}$ (approximating the lack of a resolution limit), whereas the "withering" result (blue) mimics the *Bolshoi* mass limit by only including subhaloes down to $m = \psi_{\rm res} M_0$. Lastly, "withering + disruption" (green) additionally includes the statistical treatment of artificial disruption (Section 3.2.5). The same quantities are computed from the 282 *Bolshoi* host haloes with $\log(M_0/[\,h^{-1}\mathrm{M}_\odot]) \in [14.0, 14.5]$ (black squares). When artificial disruption and withering are taken into account, `SatGen` is able to exquisitely reproduce the *Bolshoi* bias function. In the absence of such numerical limitations, `SatGen` predicts a nearly unbiased radial profile (in agreement with Han et al., 2016).

3.3.2.1 Number density and radial bias profiles

We begin by studying $\mathrm{d}\tilde{N}/\mathrm{d}x^3|_{\rm sub}$ and the corresponding bias function in Fig. 3.7. Since we aim to make direct comparisons to *Bolshoi*, we set $\psi_{\rm res} = m_{\rm res,B}/M_0$, where $m_{\rm res,B} = 10^{9.83}\ h^{-1}\mathrm{M}_\odot$ corresponds to the 50-particle halo limit that we impose on the *Bolshoi* results. For $M_0 = 10^{14.2}\ h^{-1}\mathrm{M}_\odot$, this corresponds to $\log(\psi_{\rm res}) = -4.37$. We compare the mean results obtained from 2,000 `SatGen` trees with the mean of the 282 *Bolshoi* host haloes with $\log(M_0/[\,h^{-1}\mathrm{M}_\odot]) \in [14.0, 14.5]$. The shaded regions denote the 16–84 percentiles of the halo-to-halo variance. The `SatGen` results are shown for three cases. The "fiducial" result considers all subhaloes with $m/m_{\rm acc} \geq \phi_{\rm res} = 10^{-5}$, whereas the "withering" result is limited to subhaloes with $m \geq \psi_{\rm res} M_0$. Lastly, the "withering + disruption" result includes the impact of the artificial disruption mechanism and thus can be compared directly to *Bolshoi*.

The radial profile of *Bolshoi* subhaloes becomes increasingly biased towards the central region of the host, something that has been pointed out in numerous previous studies (e.g., Diemand et al., 2004; Springel et al., 2008; Han et al., 2016). The *Bolshoi* radial profile and bias function are reproduced exquisitely by `SatGen`, but only when the impact of both withering and artificial disruption are included. Modeling the

simulation mass limit alone is sufficient to reproduce the *Bolshoi* mean curves within the halo-to-halo variance of `SatGen`; however, when artificial disruption is also taken into account, the mean curves are brought into near perfect agreement. Artificial disruption further suppresses the mean $\mathrm{d}\tilde{N}/\mathrm{d}x^3|_{\mathrm{sub}}$ by roughly a factor of two in the central region of the host. When all subhaloes can instead evolve down to $m/m_{\mathrm{acc}} = 10^{-5}$, regardless of m_{acc}, we find that the radial bias is completely eliminated. In fact, we obtain a slight overabundance of subhaloes towards the centre. This is due to dynamical friction, as we obtain a fully unbiased radial profile when we set $\beta_{\mathrm{DF}} = 0$ (i.e., no dynamical friction). We note that Han et al. (2016) report a similar finding in the *Aquarius* simulations (Springel et al., 2008). By following the most-bound particle at accretion of all subhaloes (regardless of whether or not the subhalo survives to the present day), they find a dynamical friction-driven overabundance of subhalo remnants in the halo centre that decreases towards a fully unbiased profile as m_{acc} decreases. Taken together with `SatGen`, these results demonstrate that the chief cause of the dearth of subhaloes in the central regions of haloes is the limiting mass resolution of the simulation. It is neither physical nor primarily a manifestation of artificial disruption; the latter only makes a relatively modest impact.

3.3.2.2 Projected enclosed substructure fraction

Fig. 3.8 compares the $F_{\mathrm{sub}}(< X)$ predictions of `SatGen` to the results of *Bolshoi*.[11] We use the same `SatGen` data, simulation data, and plotting conventions as in Fig. 3.7, with the only difference being that the curves/points correspond to sample *medians*. Since the (projected) enclosed substructure mass fraction, $F_{\mathrm{sub}}(< X)$, is simply a mass-weighted radial profile, and since `SatGen` reproduces both the SHMF and radial profile of *Bolshoi* subhaloes, it should come as little surprise that the model also succeeds at predicting $F_{\mathrm{sub}}(< X)$. Once again, when we include the effects of both withering and artificial disruption, the model predictions are in nearly perfect agreement with *Bolshoi*. Without accounting for artificial disruption, the median *Bolshoi* $F_{\mathrm{sub}}(< X)$ curve barely lies within the halo-to-halo variance of the withering-only prediction (for small X). Similar to $\mathrm{d}\tilde{N}/\mathrm{d}x^3|_{\mathrm{sub}}$, at $X \approx 0.1$, artificial disruption suppresses the median F_{sub} by roughly a factor of two. The difference between the fiducial and withering-only model prediction is quite small, which lies in stark contrast to the number density profile. The reason for this is that the enhanced resolution of the fiducial model predominantly results in an increased abundance of highly stripped low-m_{acc} subhaloes, which contribute little to the total substructure mass but make up a substantial portion of the number density. As we discuss in Section 3.3.3, the substructure mass fraction is primarily sensitive to the merger tree resolution

11. Since the simulation halo catalogs are constructed such that subhaloes must be instantaneously located within the virial radius of their host, we also only consider `SatGen` subhaloes within the three-dimensional virial extent of the host halo when computing $F_{\mathrm{sub}}(< X)$.

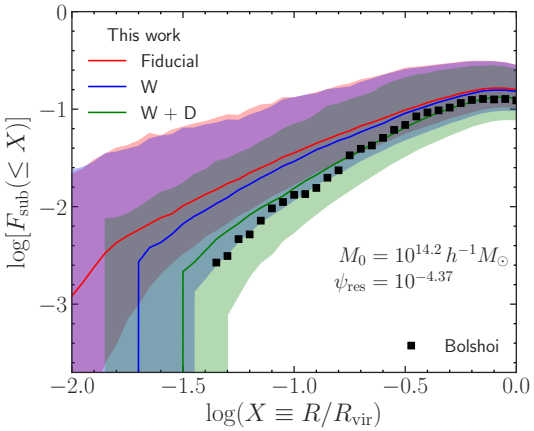

Figure 3.8: The fraction of mass enclosed within a given projected host-centric radius that is bound in *first-order* subhaloes, $F_{\rm sub}(< X)$, as defined in equation (3.20). The same `SatGen` predictions, *Bolshoi* results, and plotting conventions are used as in Fig. 3.7, with the exception being that the curves/points instead correspond to the sample medians. When both withering and artificial disruption are emulated, `SatGen` closely reproduces the *Bolshoi* $F_{\rm sub}(< X)$ profile. The substructure mass fraction is only weakly enhanced by the additional resolution in $m/m_{\rm acc}$-space afforded by the fiducial model, but it is reasonably sensitive to $\psi_{\rm res}$ (see Section 3.3.3).

($\psi_{\rm res}$).

3.3.2.3 Dependence of $d\tilde{N}/dx^3|_{\rm sub}$ on subhalo properties

We have demonstrated that by properly modeling the effects of withering and artificial disruption on the subhalo population, `SatGen` can successfully reproduce the radial distribution of simulated subhaloes. We now take a closer look at the $d\tilde{N}/dx^3|_{\rm sub}$ predictions of our fiducial model in the absence of these numerical depletion channels. Here, we analyze the results of 2,000 `SatGen` trees with $M_0 = 10^{14.2}\ h^{-1}{\rm M}_\odot$ at $z = 0$ and $\psi_{\rm res} = \phi_{\rm res} = 10^{-5}$. Thus, the lowest-$m_{\rm acc}$ subhaloes are tracked all the way down to $10^{4.2}\ h^{-1}{\rm M}_\odot$. In Fig. 3.9, we plot the mean $d\tilde{N}/dx^3|_{\rm sub}$ computed using subhaloes from these trees while varying the lower limit of several properties: (i) $m_{\rm acc}/M_0$, (ii) $m/m_{\rm acc}$, (iii) $V_{\rm peak}/V_{\rm vir,h}$, where $V_{\rm peak}$ is the peak $V_{\rm max}$ attained by the subhalo over its life (in `SatGen`, this is equivalent to the $V_{\rm max}$ at accretion, $V_{\rm max,acc}$), (iv) $V_{\rm max}/V_{\rm peak}$, (v) m/M_0, and (vi) $\log(1+z_{\rm acc})$. For comparison, we also plot the mean *Bolshoi* $d\tilde{N}/dx^3|_{\rm sub}$ computed using all subhaloes (i.e., the same as in Fig. 3.7) in each panel. Lastly, in order to facilitate a comparison with the segregation study of van den Bosch et al. (2016), we also compute the Spearman rank correlation coefficient, $r_{\rm s}$, between $r/r_{\rm vir}$ and each of the six properties computed with all subhaloes (denoted $r_{\rm s,all}$) and with subhaloes that would survive *Bolshoi* withering ($m > m_{\rm res,B}$) and artificial disruption (denoted $r_{\rm s,W+D}$, which can be directly compared to the *Bolshoi* results, $r_{\rm s,B}$).

The normalized radial profile is nearly independent of $m_{\rm acc}/M_0$ and $V_{\rm peak}/V_{\rm vir,h}$. When both withering

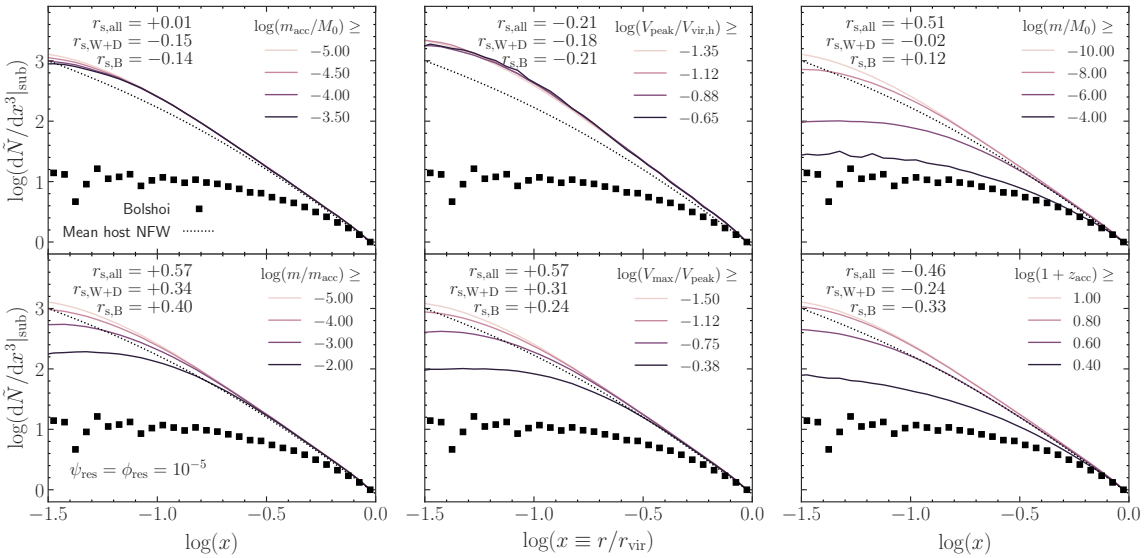

Figure 3.9: The subhalo radial profiles for 2,000 `SatGen` trees with $M_0 = 10^{14.2}\ h^{-1}\mathrm{M}_\odot$ at $z = 0$ and $\psi_{res} = \phi_{res} = 10^{-5}$. In each panel, the subhaloes are segmented by a different property and the mean $d\tilde{N}/dx^3|_{sub}$ is computed for each lower bound. In order to assess the amount of bias, we plot the mean density profile of the host in each panel (dotted line). We overplot the mean *Bolshoi* radial profile of all subhaloes in each panel (black squares). We compute the Spearman coefficient between r/r_{vir} and each property for all subhaloes ($r_{s,all}$) and the subhaloes that would survive *Bolshoi* withering and artificial disruption ($r_{s,W+D}$). There is little dependence on m_{acc}/M_0 and $V_{peak}/V_{vir,h}$. As evidenced by the m/m_{acc} and V_{max}/V_{peak} panels, highly-stripped subhaloes follow the host density profile with little bias whereas minimally-stripped systems are less commonly found in the halo centre. Similarly, massive, recently accreted subhaloes are biased towards the outer halo whereas the inclusion of older, less massive subhaloes leads to a more unbiased profile. Withering and artificial disruption tend to weaken (or reverse) the Spearman correlation between each property and r/r_{vir}, bringing our $r_{s,W+D}$ into good agreement with *Bolshoi* ($r_{s,B}$, as computed in van den Bosch et al., 2016).

and artificial disruption are taken into account, we find that r_s for each of these properties is consistent with the corresponding *Bolshoi* result reported by van den Bosch et al. (2016). The $m/m_{\rm acc}$, m/M_0, $V_{\rm max}/V_{\rm peak}$, and $\log(1+z_{\rm acc})$ panels all tell a similar story: older, less massive, and highly-stripped subhaloes follow the host potential with minimal bias. However, recently accreted, massive, and minimally-stripped systems are biased towards the halo outskirts. These trends are weakened by withering and artificial disruption, bringing the $r_{\rm s,W+D}$ for each into good agreement with van den Bosch et al. (2016).

Taken together, `SatGen` predicts that the full subhalo population should exhibit little bias with respect to the host. The dearth of subhaloes in the halo centre, which is found universally in dark matter-only simulations (e.g., Ghigna et al., 1998; Springel et al., 2001; Diemand et al., 2004; Springel et al., 2008; Han et al., 2016), is a result of inadequate resolution that causes the non-physical elimination of old, highly-stripped subhalo remnants that should be abundant in the host core.

3.3.3 Substructure mass fractions

We denote the fraction of matter bound into subhaloes within the virial radius of the host as $f_{\rm sub}(< r_{\rm vir})$. In this section, we study how the `SatGen` predictions of $f_{\rm sub}(< r_{\rm vir})$ vary with resolution limit, set by $\psi_{\rm res}$, and how they are affected by artificial disruption. In what follows, we write $f_{\rm sub}(\psi_{\rm res})$ to represent the value of $f_{\rm sub}(< r_{\rm vir})$ computed from first-order subhaloes with $m > \psi_{\rm res} M_0$. Written explicitly,

$$f_{\rm sub}(\psi_{\rm res}) = \frac{1}{M_0} \sum_{\substack{r<r_{\rm vir} \\ m>\psi_{\rm res} M_0}} m_i \,, \tag{3.21}$$

where the summation runs over first-order subhaloes only. We conclude the section by demonstrating that $f_{\rm sub}$ is insensitive to small changes in the stripping efficiency parameter, α, and the dynamical friction strength, $\beta_{\rm DF}$.

3.3.3.1 Comparison of $f_{\rm sub}(\psi_{\rm res})$ to *Bolshoi*

We begin by demonstrating the $\psi_{\rm res}$-dependence of $f_{\rm sub}$. Here, we include artificial disruption in the `SatGen` predictions in order to facilitate comparisons with *Bolshoi*. In Fig. 3.10, we plot $f_{\rm sub}(\psi_{\rm res})$ for several different halo masses. The `SatGen` predictions are obtained using 10,000 trees (with $\psi_{\rm res} = 10^{-4}$) of haloes with $M_0 = 10^{11-14}\ h^{-1}{\rm M}_\odot$ at $z = 0$. The *Bolshoi* results are computed from the 8815, 4713, 1138, and 244 host haloes with $M_0 = 10^{11\pm0.01}$, $10^{12\pm0.02}$, $10^{13\pm0.04}$, and $10^{14\pm0.1}\ h^{-1}{\rm M}_\odot$ at $z = 0$. For each halo mass, we only show results down to the $\psi_{\rm res}$ that corresponds to the 50-particle *Bolshoi* mass limit. As is evident, when combined with the *Bolshoi*-calibrated artificial disruption mechanism, `SatGen` is able to accurately reproduce the subhalo

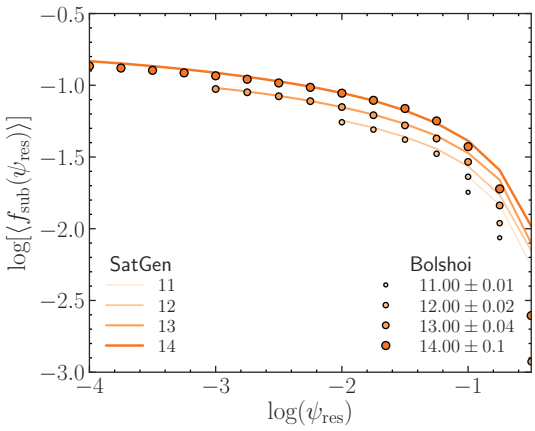

Figure 3.10: The average fraction of mass bound in subhaloes with $m > \psi_{\rm res} M_0$ within $r_{\rm vir}$ of host haloes of a given M_0, as defined in equation (3.21). The `SatGen` predictions are generated using 10,000 trees for each halo mass and are suppressed via the artificial disruption model (Section 3.2.5). The masses in the legend are reported in $\log(M_0/[\,h^{-1}{\rm M}_\odot])$. We plot the curves down to the $\psi_{\rm res}$ that corresponds to the *Bolshoi* 50-particle mass limit for each M_0. The model predictions agree well with the simulation results over a range of M_0.

statistics (and their resolution dependence) of simulated haloes over several orders of magnitude in mass. There is some tension for $\psi_{\rm res} \gtrsim 0.1$, indicating that the SHMFs of `SatGen` and *Bolshoi* disagree at the massive end. However, this likely reflects uncertainties with the (sub)halo finder used to analyze the simulation results rather than a shortcoming of `SatGen` (see van den Bosch & Jiang 2016 for a detailed discussion).

3.3.3.2 Mass-dependence and halo-to-halo variance of $f_{\rm sub}$

Fig. 3.11 plots $f_{\rm sub}(\psi_{\rm res} = 10^{-4})$ as a function of host halo mass. These results have been obtained using 10,000 trees each (with $\psi_{\rm res} = 10^{-4}$) for haloes with $\log(M_0/[\,h^{-1}{\rm M}_\odot]) \in [11, 15]$ at $z = 0$. Note that we have not included our treatment of artificial disruption here and the results are thus intended to reflect estimates of the true subhalo mass fractions in the absence of numerical artifacts. The left-hand panel shows the mean, median and 16–84 percentiles for both first- and second-order subhaloes, as indicated, whereas the right-hand panel plots the corresponding cumulative distribution functions.

Overall, the trends shown are consistent with the orbit-averaged model used by Jiang & van den Bosch (2017): $f_{\rm sub}$ increases with M_0 and the halo-to-halo variance decreases slightly with M_0. As discussed in detail in Jiang & van den Bosch (2017), this halo-to-halo variance is predominately driven by variance in the halo mass accretion histories (see also e.g., Giocoli et al., 2010; Green et al., 2020). The second-order $f_{\rm sub}$ also increases with M_0, has much larger log-scatter than the total $f_{\rm sub}$, and its mean is smaller by a factor of $\approx 15 - 30$. This difference between first- and second-order $f_{\rm sub}$ is considerably larger than predicted by Jiang & van den Bosch (2017), which is primarily due to the fact that `SatGen` allows higher-order subhaloes

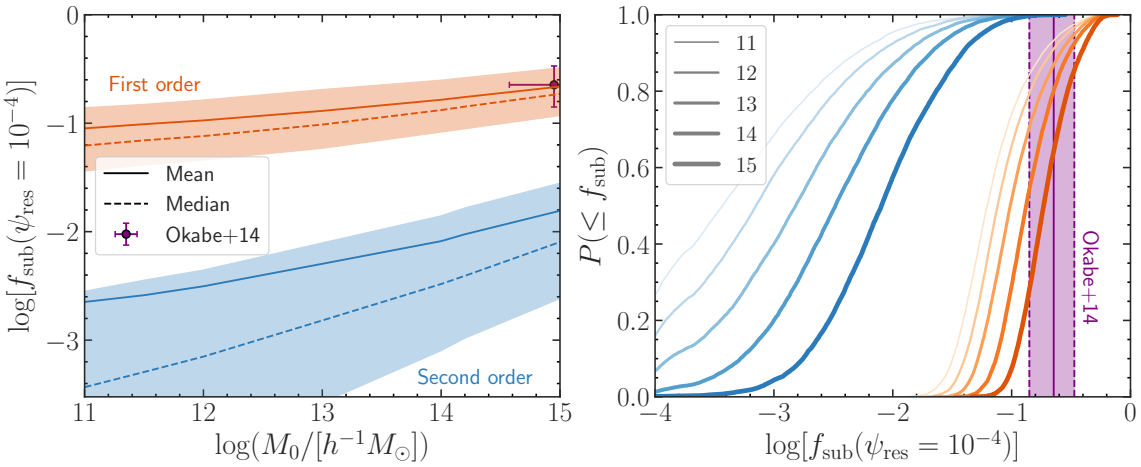

Figure 3.11: The total (first-order) and second-order $f_{\rm sub}(\psi_{\rm res}=10^{-4})$ predictions of `SatGen` in the absence of artificial disruption. The mean, median, and 16–84 percentile halo-to-halo variance of $f_{\rm sub}$ (*left*) as well as the corresponding cumulative distribution function (*right*) are computed using 10,000 trees of haloes with $M_0 = 10^{11-15}\ h^{-1}{\rm M}_\odot$ at $z = 0$. The masses in the legend are reported in $\log(M_0/[\ h^{-1}{\rm M}_\odot])$. Due to the inclusive mass definition, the first-order $f_{\rm sub}$ includes the mass of subhaloes of all-orders, whereas the second-order $f_{\rm sub}$ includes the mass of subhaloes of order-2 and higher. For comparison, we plot the gravitational lensing estimate of $f_{\rm sub}(\psi_{\rm res}=10^{-3})$ for the Coma cluster measured by Okabe et al. (2014), finding excellent agreement with our model predictions.

to be stripped from their parent subhalo (see Section 3.2.4.3).

For comparison, we also plot the result of Okabe et al. (2014), who used weak gravitational lensing to infer $f_{\rm sub} = 0.226^{+0.111}_{-0.085}$ for the Coma cluster,[12] which is assumed to have a mass of $M_0 = 8.92^{+20.05}_{-5.17}\times10^{14}\ h^{-1}{\rm M}_\odot$ (Okabe et al., 2010). Our `SatGen` predictions are in excellent agreement with this measurement, demonstrating consistency between observations and the ΛCDM paradigm.

3.3.3.3 Impact of disruption on $f_{\rm sub}$

Overall, the results of previous subsections illustrate that artificial disruption impacts subhalo statistics less significantly than the factor of two suggested by GB19. We now formalize this by comparing `SatGen` predictions of $f_{\rm sub}(\psi_{\rm res} = m_{\rm res,B}/M_0)$ with and without the impact of artificial disruption included (but with the same degree of withering in both cases since $\psi_{\rm res}$ is fixed). For this test, we use 10,000 trees with $\psi_{\rm res} = 10^{-4}$ and $M_0 = 10^{12}$ and $10^{13}\ h^{-1}{\rm M}_\odot$ as well as 2,000 trees with $\psi_{\rm res} = 10^{-5}$ and $M_0 = 10^{14.2}\ h^{-1}{\rm M}_\odot$ in order to estimate $f_{\rm sub}(\psi_{\rm res} = m_{\rm res,B}/M_0)$ with and without disruption.[13] We find that artificial disruption results in a relative suppression of $f_{\rm sub}(\psi_{\rm res} = m_{\rm res,B}/M_0)$ by 8%, 10%, and 12% for $M_0 = 10^{12}$, 10^{13}, and

12. This subhalo mass fraction is measured with $\psi_{\rm res} = 10^{-3}$, rather than 10^{-4}. As shown in Fig. 3.10, the mean $f_{\rm sub}(\psi_{\rm res}=10^{-3})$ is $\lesssim 0.1$ dex smaller than $f_{\rm sub}(\psi_{\rm res}=10^{-4})$ for high-mass host haloes, which is negligible compared to both the halo-to-halo variance and the measurement error.

13. Note that we need additional resolution for the high-mass case in order to resolve the merger trees down to $m_{\rm res,B}$.

$10^{14.2}\ h^{-1}\mathrm{M}_{\odot}$, respectively. Indeed, this level of suppression is significantly less than a factor of two (i.e., 50%). As already discussed in Section 3.3.1, the primary reason that the GB19 estimate of the artificial disruption impact is much larger is that the orbit-averaged model on which their estimate is based does not account for splashback haloes (i.e., the fact that at any moment in time about half of all haloes ever accreted by the host are located outside of the virial radius).

3.3.3.4 Insensitivity of $f_{\rm sub}$ to the model parameter choices

The substructure mass fraction is a useful summary statistic for illustrating how sensitive `SatGen` is to our model parameters (the stripping efficiency, α, and the dynamical friction strength, $\beta_{\rm DF}$). For this test, we once again focus on $M_0 = 10^{14.2}\ h^{-1}\mathrm{M}_{\odot}$ haloes. We use 10,000 trees with $\psi_{\rm res} = 10^{-4}$ and evolve the subhaloes using each of the following cases: (i) our fiducial parameters ($\beta_{\rm DF} = 0.75$ and $\alpha = \alpha(c_{\rm vir,s}/c_{\rm vir,h})$ described by equation [3.15]), (ii) fiducial $\beta_{\rm DF} = 0.75$ and fixed $\alpha = 0.6$, and (iii) fiducial $\alpha = \alpha(c_{\rm vir,s}/c_{\rm vir,h})$ and the 'natural' $\beta_{\rm DF} = 1.0$ (i.e., Chandrasekhar dynamical friction without a correction factor). As our benchmark, we consider the fractional change in the mean $f_{\rm sub}(\psi_{\rm res} = 10^{-4})$ relative to the fiducial case. Setting $\alpha = 0.6$ results in a 2% relative increase in $f_{\rm sub}$ relative to fiducial. Increasing $\beta_{\rm DF}$ from 0.75 to 1.0 results in a $\sim$ 4% relative decrease in $f_{\rm sub}$. The level of impact on other statistics (i.e., SHMF, radial profiles) is comparable. Hence, we conclude that our model predictions are reliable at the level of a few percent and that the uncertainties are small in comparison to the halo-to-halo variance. The sensitivity to these parameters is also significantly smaller than the impact of artificial disruption on the results of cosmological simulations, making `SatGen` a more reliable alternative for studying the substructure of dark matter haloes.

3.3.4 Total W + D disruption rate

The artificial disruption mechanism of Section 3.2.5 is constructed such that the D channel population of `SatGen` subhaloes has an $f_{\rm dis}$ distribution consistent with that of *Bolshoi*. However, this alone is not sufficient to guarantee that the W + D numerical disruption *rate* of `SatGen` subhaloes is in agreement with the 13%/Gyr that van den Bosch (2017) measured from the W + D channel *Bolshoi* subhaloes. In order to make a fair comparison between the `SatGen` W + D disruption rate and *Bolshoi*, we run the following test. Starting with the same sample of *Bolshoi* host halo masses (at $z \sim 0$) as used in van den Bosch (2017), we randomly sub-sample 40,000 masses from the total of ~160,000. Rather than use a fixed $\psi_{\rm res}$, we instead set $\psi_{\rm res} = m_{\rm res,B}/M_0$, where $m_{\rm res,B} = 10^{9.83}\ h^{-1}\mathrm{M}_{\odot}$ is the 50-particle *Bolshoi* resolution mass. Following the procedure of van den Bosch (2017), we determine the W + D disruption rate by measuring the fraction of the subhaloes

present at $z = 0.0174$ (i.e., 240 Myr ago) that have been disrupted (either via the W or D channel) by $z = 0$. In particular, our $z = 0.0174$ sample consists of all subhaloes that have merged with the host, have a mass above both $m_{\rm res,B}$ and the assigned $m_{\rm dis}$ (i.e., it has neither disrupted nor withered by $z = 0.0174$), and have an instantaneous orbital radius within $r_{\rm vir}$ of the host centre.

The subset of this sample with a $z = 0$ mass below either $m_{\rm res,B}$ or its assigned $m_{\rm dis}$ are counted as having numerically disrupted between $z = 0.0174$ and $z = 0$. We convert this disruption fraction into a rate by dividing it by the 240 Myr time interval considered. Using this approach, we determine that the combination of withering and our artificial disruption mechanism yields a W + D numerical disruption rate of ∼16.7%/Gyr, which is only slightly larger than the 13%/Gyr that van den Bosch (2017) measured in *Bolshoi*. Hence, we conclude that our implementation of artificial disruption in `SatGen` accurately reproduces this numerical artifact in the *Bolshoi* simulation. However, we caution that it may not adequately describe artificial disruption in other simulations, each of which is likely to have subtly different disruption statistics. The real strength of `SatGen` is not its ability to reproduce the results of cosmological simulations but rather to make reliable predictions that are free from the numerical limitations that hamper such simulations.

3.4 Summary and Discussion

This work represents the culmination of several previous studies aimed at quantifying the impact of artificial disruption on state-of-the-art dark matter-only cosmological simulations. Studying the evolution of *Bolshoi* subhaloes, van den Bosch (2017) found that the combined effect of the finite mass resolution (i.e., withering) and artificial disruption results in rapid depletion of the subhalo population. In the follow-up studies of van den Bosch et al. (2018) and van den Bosch & Ogiya (2018), the authors used a combination of analytical arguments and idealized numerical experiments to demonstrate that complete physical disruption of ΛCDM subhalo remnants is exceedingly rare, concluding that the majority of disruption seen in cosmological simulations must be numerical in nature. Following this, Ogiya et al. (2019) released the *DASH* library of high-resolution idealized simulations of halo mergers. This data release marked the beginning of a research program focused on developing a new semi-analytical model of subhalo evolution that is calibrated independently of cosmological simulations, enabling its predictions to be free of the effects of artificial disruption. Thus, GB19 used *DASH* to construct an accurate model of the evolved subhalo density profile, which is a simple function of the initial profile and the fraction of mass lost since infall (similar to the approaches of, e.g., Hayashi et al., 2003; Peñarrubia et al., 2010; Drakos et al., 2017; Errani & Navarro, 2021). Additionally, using the orbit-averaged subhalo evolution model and artificial disruption mechanism of Jiang & van den Bosch (2016), GB19 inferred that artificial disruption could potentially be responsible for

suppressing the SHMF normalization by as much as a factor of two. Recently, Jiang et al. (2021) released the `SatGen` library, a new semi-analytical modeling framework for studying subhalo and satellite galaxy evolution in a full dynamical context (i.e., the orbits of individual subhaloes are integrated instead of using an orbit-averaged approach).

In the present chapter, we used `SatGen` as a scaffolding to develop a comprehensive model of substructure evolution that is not adversely impacted by the limitations of artificial disruption and simulation resolution limits. To this end, we made several modifications and improvements to `SatGen`, which we summarize below:

- The initial orbits of infalling subhaloes are sampled using the state-of-the-art model of Li et al. (2020) (see Section 3.2.2). This model marks an improvement over previous approaches (e.g., Zentner et al., 2005a; Wetzel, 2011; Jiang et al., 2015) because it is expressed as a general function of the host halo peak height and host-to-subhalo mass ratio. Furthermore, the free parameters of the model were fit using a large simulation suite.

- The evolved subhalo density profiles (ESHDPs) are characterized using the model of GB19 (Section 3.2.4.1). At infall, subhaloes are assumed to have NFW profiles. However, as mass is stripped and $m/m_{\rm acc}$ decreases, the profile becomes tidally truncated in a manner consistent with the evolution of *DASH* subhaloes.

- In line with the original `SatGen` implementation, the instantaneous subhalo mass-loss rate (Section 3.2.4.2) is written according to equation (3.11), which depends on the King (1962) tidal radius (computed using the ESHDPs), the local dynamical time, and the "stripping efficiency", α. We recalibrated α (equation [3.15]) so that the mass-loss model accurately reproduces the $m(t)/m_{\rm acc}$ trajectories of *DASH* subhaloes.

- The strength of the (Chandrasekhar) dynamical friction is controlled by a correction factor, $\beta_{\rm DF}$, which we calibrate such that `SatGen` reproduces the $m_{\rm acc}/M_0$-dependence of the $\langle r/r_{\rm vir}\rangle - z_{\rm acc}$ relation of *Bolshoi* subhaloes (Section 3.2.6). We have demonstrated that the resulting best-fit value ($\beta_{\rm DF} = 0.75$) is not adversely affected by artificial disruption in the *Bolshoi* simulation.

- In order to assess the impact of artificial disruption on simulations, we developed a model that reproduces the statistical properties of disruption in *Bolshoi* that can be optionally applied to `SatGen` results. We found that the $f_{\rm dis}$ distribution of disrupted (D channel) *Bolshoi* subhaloes is well-described by a family of log-normal distributions, the parameters of which are functions of $m_{\rm acc}$ (Section 3.2.5).

- `SatGen` is ideally suited to assess the impact of the resolution limit of numerical simulations by only including subhaloes with a final mass that lies above the merger tree resolution (i.e., $m > \psi_{\rm res} M_0$). In addition, by instead allowing each subhalo to evolve down to arbitrary $\phi_{\rm res} = m/m_{\rm acc}$ (here, we have used values as low as $\phi_{\rm res} = 10^{-5}$), `SatGen` can model the subhalo population with an effectively "arbitrary resolution".

We used this updated model to predict subhalo mass and maximum circular velocity functions, number density profiles, radial bias profiles, and substructure mass fractions. We considered the effect of both the simulation mass limit and artificial disruption on each quantity and studied the dependence of $f_{\rm sub}$ on host halo mass. We summarize our most notable findings below:

- When the effects of both withering and artificial disruption are included, `SatGen` yields subhalo demographics in excellent agreement with *Bolshoi*.

- Artificial disruption only results in a $\sim 8-12\%$ suppression of $f_{\rm sub}(< r_{\rm vir})$ and a $\sim 20\%$ suppression of the SHMF. While still significant, this greatly ameliorates previous concerns that the overall abundance of dark matter subhaloes could be artificially suppressed by a factor of two. However, the impact of artificial disruption is more pronounced at smaller host-centric radii, where it halves both $F_{\rm sub}(< X)$ and $\mathrm{d}\tilde{N}/\mathrm{d}x^3|_{\rm sub}$ within $\sim 0.1 r_{\rm vir}$.

- By comparing the SHMF computed by including only subhaloes within $r_{\rm vir}$ (i.e., consistent with simulation approaches) to the SHMF computed by including *all* surviving subhaloes ever accreted by the host, we infer that splashback haloes make up roughly half of the total subhalo population. This is in good agreement with results from several simulation studies (e.g., Gill et al., 2004; Ludlow et al., 2009; Bakels et al., 2021). Hence, it is essential that semi-analytical models of subhalo and satellite galaxy evolution properly account for the splashback population. This is naturally achieved with full dynamical models, such as `SatGen`, which integrate the orbits of individual subhaloes. At the same time, it indicates a serious limitation of orbit-averaged approaches, such as those used in van den Bosch et al. (2005b) and Jiang & van den Bosch (2016).

- We have demonstrated that the radial bias in the subhalo number density (i.e., the dearth of subhaloes in the halo centre relative to the host density profile), a feature that is consistently present in dark matter-only simulations (e.g., Ghigna et al., 1998; Springel et al., 2001; Diemand et al., 2004; Springel et al., 2008; Han et al., 2016), is predominantly an artifact of the simulation mass resolution (at least in the absence of baryonic processes) and not of artificial disruption. The latter only slightly enhances the bias and is subdominant to the impact of the mass resolution. By allowing subhaloes to evolve

down to arbitrarily low $m/m_{\rm acc}$ (as opposed to having a fixed absolute mass limit), the radial bias is completely eliminated. In fact, dynamical friction causes a slight enhancement of the subhalo number density relative to the host profile near the halo centre, which marks a complete reversal of the trend seen in simulations.

Although the model presented here is able to accurately reproduce the subhalo statistics of a cosmological simulation when its numerical limitations are properly taken into account, the true strength of the updated version of `SatGen` presented here lies in the fact that it can be used to predict subhalo demographics with an arbitrarily high resolution and in the absence of artificial disruption. We have therefore made the updated code publicly available in the hope that it will enable/accommodate a wide variety of future research programs. For example, `SatGen` could prove a powerful tool to investigate claimed discrepancies between simulations and observations regarding the abundance and central concentration of dark matter substructure (e.g., Carlsten et al., 2020; Meneghetti et al., 2020) and/or the dark matter deficiency of associated satellite galaxies (e.g., Ogiya, 2018; Jackson et al., 2021).

Chapter 4

Impact of the galactic disc

This chapter has been accepted for publication as an article by Green et al. (2021b) in the Monthly Notices of the Royal Astronomical Society by Oxford University Press and was in press at the time of completion of this dissertation.

4.1 Background

THE substructure present in dark matter (DM) haloes is the outcome of a hierarchical assembly process combined with tidal and impulsive forces that work to dissolve it. Since the particle nature of DM impacts the mass function and density profiles of DM haloes, it also affects the demographics of its substructure. For example, if DM is "warm", the abundance of low-mass subhaloes is suppressed relative to that which is predicted for cold dark matter (e.g., Lovell et al., 2014; Bose et al., 2017). If DM undergoes significant self-interaction, or is an ultra-light boson, the inner halo density profile becomes cored (e.g., Kaplinghat et al., 2016; Robles et al., 2017; Burkert, 2020) and is less resilient to tidal forces (Peñarrubia et al., 2010), resulting in an overall suppression of substructure. This powerful potential to place constraints on DM has prompted various observational attempts to quantify the abundance of DM substructure, including searches for gaps in stellar streams (e.g., Erkal et al., 2016; Banik et al., 2019; Bonaca et al., 2020), measurements of gravitational lensing distortions (e.g., Vegetti et al., 2014; Hezaveh et al., 2016; Nierenberg et al., 2017), indirect detection studies that search for DM annihilation signals (e.g., Stref & Lavalle, 2017; Somalwar et al., 2021), and measurements of the abundance of satellite galaxies (via the galaxy–halo connection; e.g., Nadler et al., 2021).

In order to fully leverage these observations to constrain DM microphysics, it is prudent that we are able to accurately predict subhalo abundances for the different DM models. Arguably, the best way to account for

all the relevant, strongly non-linear physical processes is to use full cosmo-hydrodynamical simulations of galaxy formation (e.g., Wetzel et al., 2016; Pillepich et al., 2018) at a resolution sufficient to resolve substructure in the relevant mass range. Unfortunately, the computationally demanding nature of such simulations as well as the uncertainties related to sub-grid physics modeling represent significant roadblocks for their use in such a task. As a consequence, DM-only cosmological simulations (e.g., Springel et al., 2008; Klypin et al., 2011) are typically used as a less expensive alternative. However, these simulation-based approaches are still adversely impacted by artificial subhalo disruption and limited mass resolution (van den Bosch, 2017; van den Bosch & Ogiya, 2018; Green et al., 2021a). Semi-analytical models (SAMs; e.g., Taylor & Babul, 2004; Zentner et al., 2005a; Jiang & van den Bosch, 2016; Jiang et al., 2021) provide attractive alternatives for predicting the substructure abundance in a manner that is both computationally efficient and insensitive to the particular numerical limitations of N-body simulations.

Unfortunately, DM-only simulations and SAMs typically do not account for the impact of baryons on the subhalo population, which can be quite important. For example, the aforementioned observational probes are most sensitive to the inner halo, where the central galaxy significantly influences the host potential. Several studies have demonstrated that a central galactic disc suppresses the overall subhalo abundance. For example, D'Onghia et al. (2010) grew an analytical disc potential in a high-resolution cosmological zoom-in simulation of a Milky Way-like (MW) halo and showed that substructure in the inner regions of the halo is efficiently destroyed, which they ascribed to disc shocking. More recently, Garrison-Kimmel et al. (2017) found that the suppression in subhalo abundance seen in a full physics simulation relative to a DM-only realization of the same halo can be reproduced by simply embedding a disc potential within the DM-only halo. Both Peñarrubia et al. (2010) and Errani et al. (2017) used idealized simulations to examine the impact of a central disc on the abundance of subhaloes, confirming once more that the presence of a disc can significantly deplete the subhalo population, especially towards the centre of the halo.

Since a disc potential drives additional subhalo mass loss, its presence must be properly accounted for in any successful substructure modeling endeavor. However, to date, no study has been able to assess the impact of the disc on subhalo populations in a statistically meaningful way. Recently, we introduced `SatGen` (Jiang et al., 2021), a SAM framework that can rapidly generate random substructure realizations, thereby enabling a robust treatment of the halo-to-halo variance. As shown in Jiang & van den Bosch (2017), this variance can be very large and is strongly correlated with the formation time and concentration of the host halo (see also Zentner et al., 2005a; Giocoli et al., 2010). Furthermore, `SatGen` can be used to isolate the influence of the disc from assembly history variation by studying how subhaloes from the same merger tree evolve under different host potentials. Due to its speed, `SatGen` is also ideal for assessing how sensitive the subhalo statistics are to parameters of the disc model via sweeps of the parameter space.

In this chapter, we use `SatGen` to investigate the differential impact of a galactic disc potential on the subhalo populations of Milky Way-like haloes. We initially explored the influence of a disc in Jiang et al. (2021) — here, we build upon this pilot study by greatly boosting the size of our halo sample, exploring a wide range of disc models, and incorporating a more sophisticated subhalo tidal evolution model. While our findings are in good agreement with the simulation results of Errani et al. (2017) and Garrison-Kimmel et al. (2017), our ability to study a large halo sample and, thus, probe the halo-to-halo variance sheds new light on the statistical relevance of these results. We track individual subhaloes and illustrate how their masses are altered due to an embedded central disc. We also search for the presence of a disc-driven angular bias in the spatial distribution of subhaloes, as well as show that the overall subhalo abundance is relatively insensitive to the size and growth history of the disc and is only affected by the disc mass. This manuscript is organized as follows. In Section 4.2, we first provide an overview of our semi-analytical modeling framework. The results are presented in Section 4.4, which is followed by a detailed discussion (Section 4.5) as to whether "disc shocking" or enhanced tidal stripping serves as the dominant disc-driven subhalo depletion mechanism. Finally, in Section 4.6, we summarize our findings and motivate future work.

Throughout this work, the halo mass is defined as the mass enclosed within the virial radius, $r_{\rm vir}$, inside of which the mean density is equal to $\Delta_{\rm vir}(z)$ times the critical density. For the Λ cold dark matter (ΛCDM) cosmology that we adopt ($h = 0.7$, $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_{\rm b} = 0.0465$, $\sigma_8 = 0.8$, $n_{\rm s} = 1.0$), $\Delta_{\rm vir}(z = 0) \approx 100$ and is otherwise well-described by the fitting formula presented by Bryan & Norman (1998). Throughout, we use m and M to denote subhalo and host halo masses, respectively. We use l and r to reference subhalo- and host halo-centric radii, respectively. The base-10 logarithm is denoted by log and the natural logarithm is denoted by ln.

4.2 Semi-analytical methods

This study employs the `SatGen` semi-analytical modeling framework that is presented by Jiang et al. (2021). In particular, we use the model of subhalo tidal evolution recently developed by Green & van den Bosch (2019) and Green et al. (2021a). This model has been calibrated using the *Dynamical Aspects of Subhaloes* idealized simulation library (hereafter *DASH*; Ogiya et al., 2019) to accurately reproduce the bound mass and density profiles of simulated N-body subhaloes as they orbit within an analytical host potential. We refer the reader to these papers for a comprehensive description of the model. In short, `SatGen` combines prescriptions for (i) analytical halo merger trees (Parkinson et al., 2008), (ii) subhalo orbit initialization (Li et al., 2020), (iii) orbit integration, including dynamical friction (Chandrasekhar, 1943), (iv) density profile evolution (Green & van den Bosch, 2019), and (v) tidal mass-loss (Green et al., 2021a) in order to generate subhalo

catalogs (which include both mass and position information) for ensembles of host halo realizations.

Both the host halo and the initial subhaloes (i.e., at infall) are assumed to have Navarro et al. (1997, hereafter NFW) density profiles with concentrations computed via the model of Zhao et al. (2009). Each subhalo is integrated along its orbit and experiences tidal mass loss, which is given by

$$\frac{\Delta m}{\Delta t} = -\alpha \frac{m(> l_{\rm t})}{t_{\rm char}} \,. \tag{4.1}$$

Here, $t_{\rm char}$ is the characteristic orbital time of the subhalo, $m(> l)$ is the subhalo mass that lies outside of radius l, with $l_{\rm t}$ denoting the instantaneous tidal radius (defined below), and α is a calibrated parameter that controls the stripping efficiency. Motivated by the work of Hayashi et al. (2003) and Peñarrubia et al. (2008), the density profiles of stripped subhaloes are modelled according to

$$\rho(l,t) = H(l|\, f_{\rm b}(t), c_{\rm vir,s})\, \rho(l, t_{\rm acc}), \tag{4.2}$$

where $f_{\rm b}(t)$ is the bound mass fraction of the subhalo, $c_{\rm vir,s}$ is the concentration of the subhalo *at accretion*, and $t_{\rm acc}$ denotes the time of accretion. For the 'transfer function', $H(l)$, we use the model of Green & van den Bosch (2019), which has been carefully calibrated against the *DASH* simulations.

The galactic disc, when included, is positioned at the centre of the host halo and modeled with the axisymmetric Miyamoto & Nagai (1975, hereafter MN) density profile, which has three parameters: (i) the radial scale length, $a_{\rm d}$, (ii) vertical scale height, $b_{\rm d}$, and (iii) mass, $M_{\rm d}$. We write $a_{\rm d} = f_a [M_{\rm vir}(z)/M_0]^{\beta_a} r_{\rm vir,0}$, $M_{\rm d} = f_M [M_{\rm vir}(z)/M_0]^{\beta_M} M_0$, and set $b_{\rm d}$ to be a fixed fraction of $a_{\rm d}$. Here, $M_{\rm vir}(z)$ is the mass accretion history of the host halo, $M_0 = M_{\rm vir}(z=0)$, and $r_{\rm vir,0}$ is the virial radius of the host at $z = 0$. In this work, we restrict ourselves to host haloes that reach a virial mass of $M_0 = 10^{12}\, h^{-1} M_\odot$ at $z = 0$, which corresponds to $r_{\rm vir,0} \approx 290$ kpc. Our fiducial, Milky Way-like disc is described by $f_a = 0.0125$, $b_{\rm d}/a_{\rm d} = 0.08$, $\beta_a = 1/3$, $f_M = 0.05$, and $\beta_M = 1$, such that the disc mass grows linearly with $M_{\rm vir}(z)$ and the scale length grows linearly with $r_{\rm vir}(z)$, in good agreement with both empirical constraints (Kravtsov, 2013) and simulation results (Jiang et al., 2019a). The disc properties at $z = 0$ are $a_{\rm d} \approx 3.6$ kpc, $b_{\rm d} \approx 0.3$ kpc (i.e., a relatively thin disc), and $M_{\rm d} = 5 \times 10^{10}\, h^{-1} M_\odot$, reminiscent of the Milky Way. The parametrization chosen is sufficiently flexible to enable us to study the impact of the disc growth history and structure on the subhalo population by simply varying $f_a, \beta_a, f_M, \beta_M$, and $b_{\rm d}/a_{\rm d}$. As we show in Section 4.4.5, though, the results are relatively insensitive to all but the final disc mass (f_M). We emphasize that the total mass of the halo–disc system enclosed within $r_{\rm vir}(z)$ sums to $M_{\rm vir}(z)$. In order to achieve this, we simply multiply the host NFW density profile by $[M_{\rm vir}(z) - M_{\rm d}(z)]/M_{\rm vir}(z)$.

The merger tree resolution, which sets the lower limit on the subhalo accretion mass, corresponds to $m_{\rm res} \geq 10^8\,h^{-1}M_\odot$. All subhaloes are evolved until their mass falls below $m = 10^{-5}m_{\rm acc}$. All results are averaged over 10,000 host halo merger tree realizations. Specifically, we use the same set of merger trees throughout in order to isolate the differential impact of the disc from the effect of assembly history variation.

In order to strengthen the performance of our mass-loss model with composite halo–disc hosts, we make two changes to `SatGen` relative to its specification in Jiang et al. (2021) and Green et al. (2021a). Since we are working with an axisymmetric potential, we make the following substitution for the tidal radius definition:

$$l_{\rm t} = \left[\frac{m(<l_{\rm t})/M(<r)}{2 + \frac{\Omega^2 r^3}{GM(<r)} - \frac{{\rm d}\ln M}{{\rm d}\ln r}|_r}\right]^{1/3} \Rightarrow l_{\rm t} = \left[\frac{Gm(<l_{\rm t})}{\Omega^2 - \frac{{\rm d}^2\Phi}{{\rm d}r^2}|_{\boldsymbol{r}}}\right]^{1/3}. \tag{4.3}$$

Here, $\boldsymbol{\Omega}$ is the instantaneous angular velocity of the subhalo and Φ is the gravitational potential of the host halo–disc system. Note that these two definitions of $l_{\rm t}$, introduced by King (1962), are identical when the host potential is spherically symmetric. However, the definition on the right is more general, as it does not depend on spherical averages of the host properties. To wit, the radial derivative of an axisymmetric potential is $\frac{{\rm d}\Phi}{{\rm d}r} = \frac{\partial\Phi}{\partial R}\frac{{\rm d}R}{{\rm d}r} + \frac{\partial\Phi}{\partial z}\frac{{\rm d}z}{{\rm d}r}$. Hence, for spherical hosts, the mass-loss model of equation (4.1), which was calibrated using the definition on the left, remains unchanged after this substitution. The second change is with regards to the mass-loss coefficient, α, which is a function of the ratio between the host and subhalo concentrations (see Green et al., 2021a). In order to account for the modified mass distribution in the presence of a disc, we modify the host concentration used to compute α according to the following procedure. We define a new scale radius, $r_{\rm s,d}$, as the radius within which the enclosed mass of the combined halo–disc system is the same as that which is enclosed within the scale radius of the NFW halo-only host. The modified host concentration is simply $r_{\rm vir}/r_{\rm s,d}$, which is somewhat larger than the halo-only concentration due to the compactness of the disc. Note that, unlike Jiang et al. (2021), we omit the effect of adiabatic contraction on the host concentration since we are interested in studying *relative* disc-driven subhalo depletion. However, we emphasize that a baryon-driven increase in the DM concentration is degenerate with an increased disc mass, which we demonstrate in Section 4.4.6.

Since we calibrated our mass-loss model using the *DASH* simulations, which only include spherical NFW host haloes, we must verify that the prescription remains valid for subhaloes that evolve in the combined presence of a NFW halo and a MN disc. To this end, we run an additional set of idealized simulations and compare the predictions of our mass-loss model (with the substituted $l_{\rm t}$ definition and modified host concentration definition) to the mass trajectories of the simulated subhaloes. The simulation methods and model comparison results are described in Section 4.3. In summary, we find that, after making the two

aforementioned modifications, the model generalizes well to combined halo–disc potentials and, thus, we proceed to use this modified version of `SatGen` in this work.

4.3 Idealized simulations

In order to verify that the Green et al. (2021a) mass-loss model, which is given by equation (4.1), can be used to accurately describe subhalo evolution in a combined halo–disc host system, we run a set of idealized simulations that serve as the ground truth for comparison to our model predictions. We use the same procedure as used for the *DASH* simulations (Ogiya et al., 2019). In particular, all simulations are run using the N-body code `OTOO+` (Ogiya et al., 2013), which is a GPU-accelerated tree code. The simulations follow the evolution of a live N-body subhalo (initially a NFW halo composed of $N = 10^6$ particles) as it orbits within a static NFW host halo potential (with initial sub-to-host mass ratio of $m/M = 10^{-3}$) for 36 Gyr.

In contrast to *DASH*, the host system in our test suite of simulations is composed of a NFW halo with an embedded MN disc. The host halo has a concentration of $c_{\rm vir,h} = 10$ and the subhalo has an initial concentration of $c_{\rm vir,s} = 20$, which is consistent with a typical minor merger with a Milky Way-like host. For simplicity, we only consider orbits with orbital energy equal to that of a circular orbit at the virial radius of the host, which coincides with the peak of the orbital energy distribution of infalling subhaloes in cosmological simulations (e.g., Li et al., 2020).

When constructing our simulation suite, we vary several parameters, which control properties of the disc, the inclination of the initial subhalo orbit with respect to the disc, and the radius of orbital pericentre. The fraction of the host mass in the disc is set by $f_M \equiv M_{\rm d}/M_{\rm vir} \in [0.0, 0.02, 0.05, 0.1]$. In order to preserve the total mass enclosed within $r_{\rm vir}$, we multiply the host halo density profile by $1 - f_M$. The disc scale length is varied over $f_a \equiv a_{\rm d}/r_{\rm vir} \in [0.007, 0.0125, 0.025]$, with $f_a = 0.0125$ corresponding to $a_{\rm d} \approx 3.5$ kpc for the Milky Way-mass host, while the disc scale height is controlled by varying $b_{\rm d}/a_{\rm d} \in [0.02, 0.06, 0.2]$. We consider seven orbital inclinations with $i \in [0, 15, 30, 45, 60, 75, 90]$ degrees, where i is the angle between the orbital plane and the disc (i.e., $i = 0°$ results in a subhalo that orbits in the plane of the disc). The final parameter is the orbital angular momentum, which we adjust such that the pericentric radius, $r_{\rm p}$, of the subhalo orbit *in the no-disc configuration* is equal to 15, 25, or 50 kpc.

Given the orbital energy, angular momentum, and inclination angle, we initialize the subhalo at its apocentre. Note that we use the same initial position and velocity regardless of the disc properties. Hence, the true $r_{\rm p}$ attained by the subhalo varies slightly with f_M, f_a, and $b_{\rm d}/a_{\rm d}$.

Thus, our test suite spans f_M, f_a, $b_{\rm d}/a_{\rm d}$, i, and $r_{\rm p}$. We follow the procedure laid out in Section 2.3.2 of Green et al. (2021a) to generate mass-loss model predictions and make comparisons to the bound-mass

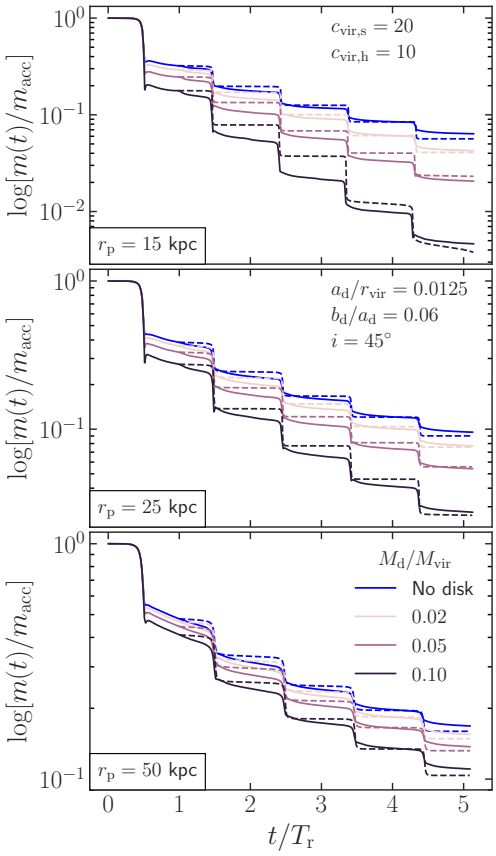

Figure 4.1: A comparison between the Green et al. (2021a) mass-loss model predictions (dashed lines) and the $m(t)/m_{\rm acc}$ trajectories of several simulated subhaloes (solid lines). The times are normalized by the radial orbital period, T_r. The host and subhalo concentrations, disc shape, orbital inclination, and orbital energy are held fixed. The model performs well over the full range of disc masses and down to small pericentric radii ($r_{\rm p}$).

trajectories, $m(t)/m_{\rm acc}$, of the simulated subhaloes. The performance of the model is illustrated in Fig. 4.1, which compares the simulation and model results for the case of a fiducial MW-like disc shape and an inclined subhalo orbit with $i = 45°$. Clearly, the model accurately captures the subhalo mass evolution for all $M_{\rm d}/M_{\rm vir}$ and $r_{\rm p}$ considered. When averaged over the full test suite, we find that our model remains unbiased with relatively low scatter in the log-residuals (less than a factor of two larger than that of the halo-only case) for longer than a Hubble time, indicating that mass evolution error will be subdominant to the halo-to-halo variance. Hence, the mass-loss model successfully captures the additional loss of mass due to the presence of a disc, validating its use in `SatGen` for the present study.

4.4 Results

4.4.1 Subhalo mass functions

We begin by assessing the impact of the disc on the abundance of subhaloes as a function of their mass. Fig. 4.2 plots the cumulative subhalo mass functions (SHMF), $N(>m/M_0)$, for the disc-less case (blue lines) as well as for several disc configurations, each with our fiducial f_a, β_a, and β_M, but with different values of the disc mass fraction, f_M, as indicated. The range of disc mass fractions covered ($f_M \in [0.025, 0.05, 0.075, 0.1]$) is motivated by estimates of the stellar mass–halo mass relation (see e.g., Moster et al., 2010; More et al., 2011; Behroozi et al., 2019). The dark and light shaded regions denote the $16-84$ and $2.5-97.5$ percentile intervals, respectively, of the individual SHMFs in the halo-only (i.e., no-disc) case, highlighting the typical halo-to-halo variance. The SHMFs in the left panel include subhaloes of all orders[1] that have an instantaneous host-centric radius at $z=0$ of $r < r_{\rm vir,0}$. In order to emphasize the pronounced effect of the disc in the halo centre, the SHMFs in the right panel are restricted to subhaloes with $r < 50$ kpc. Clearly, the disc results in a suppression of the subhalo abundance that is proportional to the disc mass fraction, f_M. However, the halo-to-halo variance in the SHMF is dramatically larger than the difference between the mean SHMFs from the various disc configurations (this finding is also present throughout the results of Jiang et al., 2021). The most massive disc, with $f_M = 0.1$, results in a ∼0.09 dex (18%) suppression in the $r < r_{\rm vir,0}$ SHMF at the low-mass end, which decreases slightly with increasing m/M_0. Our results are in excellent agreement with the idealized simulations of Errani et al. (2017), who report a 20% suppression in the SHMF of a cuspy Milky Way-mass halo due to the presence of a disc with $f_M = 0.1$ (note that they also find the effect to be reduced at the high-mass end of the SHMF). The factor of suppression due to the disc is greatly increased when we restrict the SHMF to subhaloes with $r < 50$ kpc. For example, the $f_M = 0.1$ disc drives a ∼0.25 dex (∼44%) decrease in subhaloes within 50 kpc, which is consistent with D'Onghia et al. (2010). Our mean results are also in excellent agreement with Jiang et al. (2021), indicating an overall insensitivity to our differing subhalo tidal evolution models.[2]

We find that the slope of the $r < r_{\rm vir,0}$ SHMF changes very little with f_M, ranging from -0.91 in the disc-less case to -0.89 in the $f_M = 0.1$ configuration. The same is true of the $r < 50$ kpc SHMF, which has a slope of -0.81 in the disc-less case and -0.79 when $f_M = 0.1$. Taken together, these results imply that mass segregation (or the lack thereof, cf. van den Bosch et al., 2016) is not greatly impacted by the disc. However,

1. A subhalo of order n is hosted by a (sub)halo of order $n-1$, with host haloes corresponding to order 0.

2. Note that Jiang et al. (2021) also account for the fact that the subhalo density profile may be affected by baryons prior to infall. However, this only affects the small fraction of subhaloes that host bright satellites, the analysis of which is beyond the scope of this study.

because of the slight f_M-dependence of the slope, the $r < r_{\rm vir,0}$ SHMF residuals exhibit a small amount of mass-dependence, especially for the larger f_M. In order to gauge how this plays out at the low mass end (i.e., $\log(m/M_0) < -4$), we generated 2,000 merger trees with enhanced resolution ($m_{\rm acc} \geq 10^7\,h^{-1}M_\odot$) and evolved the subhaloes using each of the five configurations introduced in Fig. 4.2. We report that the residuals for $10^{-5} \leq m/M_0 \leq 10^{-4}$ flatten off and are consistent with their values at $m/M_0 = 10^{-4}$. This convergence of the slopes in the low-m/M_0 limit indicates that the additional dynamical friction due to the disc, which only impacts more massive subhaloes, is the most likely cause of the minor f_M-dependence of the SHMF slope.

In order to aid our evaluation of the significance of "disc shocking", we consider a case where the MN disc is replaced by a spherical component with a nearly equivalent spherically enclosed mass profile. Specifically, we fit the $M(<r)$ of an Einasto (1965) profile to the $M(<r)$ of a MN disc with $b_{\rm d}/a_{\rm d} = 0.08$. Using the notation of Jiang et al. (2021, Appendix A3), the parameters of the resultant Einasto (1965) halo are $M_{\rm tot} = M_{\rm d}$ (i.e., the total mass of the system is unchanged), $n = 2.13$, and $c_2 = 43.7$. The concentration is defined with respect to the $z = 0$ virial mass definition. Using this convention, the mass and size of the spherical substitute grow identically to that of the disc by simply holding $M_{\rm tot} = M_{\rm d}$. The Einasto (1965) sphere is slightly more centrally concentrated — its enclosed mass is ∼12% larger than that of the disc at $r \approx 6a_{\rm d}$, with the $M(<r)$ of the two systems converging as r increases. We evolve the subhaloes in this composite halo–sphere host with $f_M = 0.1$, presenting the resulting SHMF as dashed lines in Fig. 4.2. The results agree quite well with those of the $f_M = 0.1$ disc. The slightly increased central concentration of the spherical replacement appears to drive a minor increase in the overall mass loss relative to the disc. This agreement between the impact of a disc and a spherical replacement is also reported by Garrison-Kimmel et al. (2017). We elaborate on the implications of this finding in Section 4.5.

4.4.2 Radial profiles

In Fig. 4.3, we shift our attention to the cumulative radial subhalo abundance profile, $N(<r/r_{\rm vir,0})$. Here, we restrict ourselves to two different subsets of the subhalo population. In the left panel, we only count subhaloes that have a maximum circular velocity at accretion, $V_{\rm max,acc}$, that is greater than 30 km/s, which roughly captures the population of subhaloes that could themselves host galaxies. We note that all such subhaloes with $m/m_{\rm acc} > 10^{-5}$ are included; however, we find that the results presented throughout this work are qualitatively insensitive to the choice of $m/m_{\rm acc}$ used as the cut-off for inclusion. In the right panel, we instead include all subhaloes with instantaneous mass at $z = 0$ of $m > 10^8\,h^{-1}M_\odot$. Once again, the dark and light shaded regions indicate the 68% and 95% halo-to-halo variance intervals in the no-disc case. The

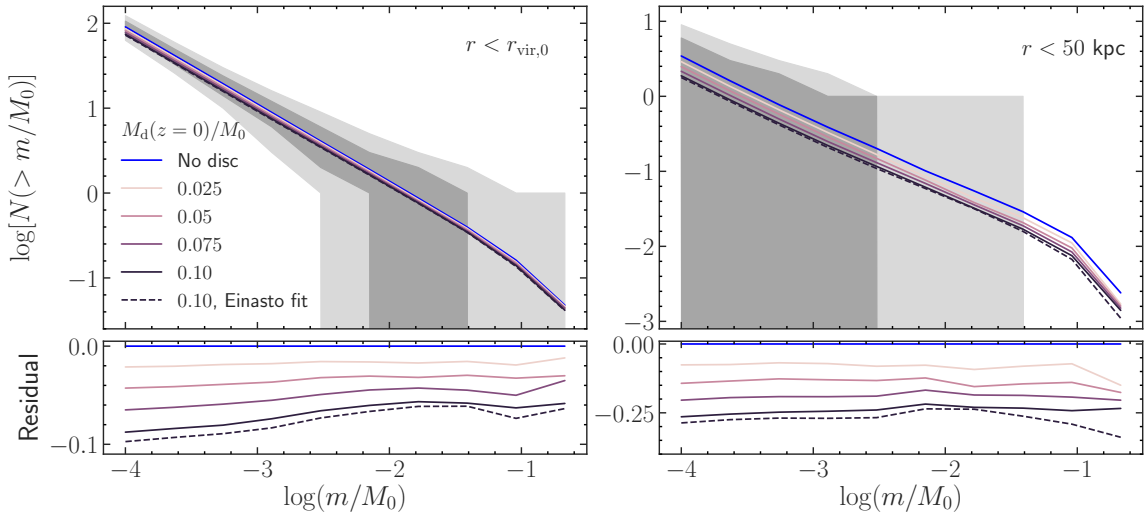

Figure 4.2: The subhalo mass function, which includes subhaloes of all orders with a $z = 0$ host-centric radius of (*left*) $r < r_{\rm vir,0}$ or (*right*) $r < 50$ kpc. The curves denote the mean SHMF taken over 10,000 trees whereas the dark and light shaded regions correspond to the $16-84$ and $2.5-97.5$ percentiles, respectively, of the individual (halo-only) trees. The different curves illustrate the dependence of the final disc mass on the subhalo population. The suppression of the subhalo abundance is proportional to the disc mass and is much larger in the halo centre. The dashed lines correspond to the replacement of the $f_M = 0.1$ disc with an Einasto (1965) sphere that has a nearly equivalent enclosed mass profile. The agreement between these dashed lines and the corresponding $f_M = 0.1$ disc curves demonstrates the insignificance of "disc shocking".

mean profile of each disc configuration is again enclosed within the halo-to-halo variance of the disc-less profile. The radial profiles further illustrate the enhanced impact of the disc on subhalo statistics towards the halo centre, as the mean profiles are increasingly suppressed with decreasing $r/r_{\rm vir,0}$. This effect is strongest on the population with $V_{\rm max,acc} > 30$ km/s, which is reduced by roughly 0.7 dex ($\approx 80\%$) within 20 kpc of the halo centre when $f_M = 0.1$. In comparison to all subhaloes with $m > 10^8\,h^{-1}M_\odot$, the population with $V_{\rm max,acc} > 30$ km/s is composed of a larger fraction of initially massive subhaloes, which experience stronger dynamical friction and, thus, enhanced tidal stripping. We find that the fractional impact of the disc on the radial profiles of the two simulated systems studied by Garrison-Kimmel et al. (2017) is within the halo-to-halo variance of our log-residuals (see also Jiang et al., 2021). Once again, the replacement of the disc with a nearly equivalent Einasto (1965) sphere results in a radial profile that agrees exquisitely well with that of the $f_M = 0.1$ disc.

4.4.3 Enhanced tidal stripping

By using the same merger tree realizations for all of the halo–disc configurations, `SatGen` enables us to directly assess the impact of the disc on *individual* subhaloes. In Fig. 4.4, we use a log-density heatmap to

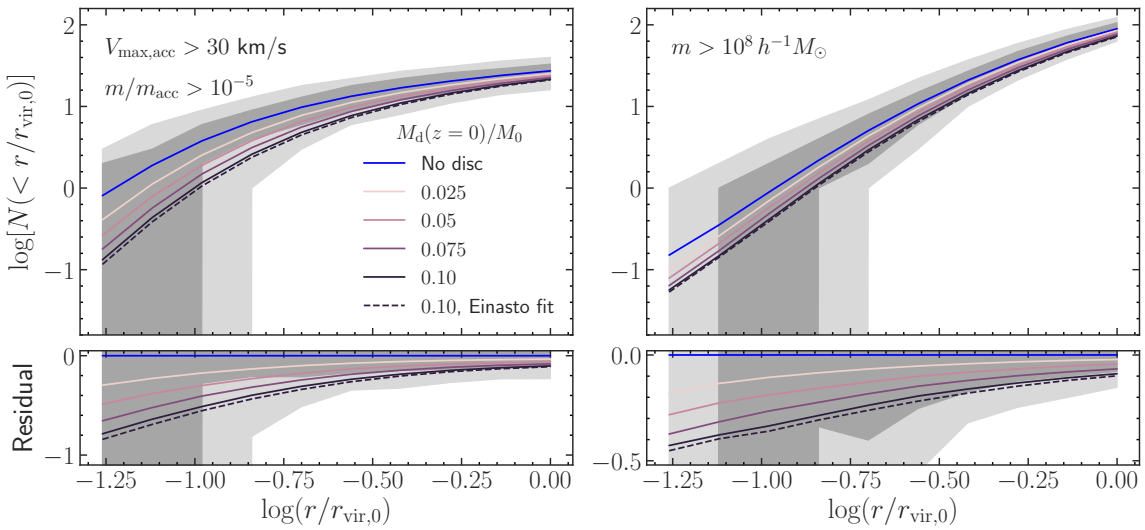

Figure 4.3: The cumulative radial subhalo abundance profile, which includes subhaloes of all orders with (*left*) $V_{\rm max,acc} > 30$ km/s and $m/m_{\rm acc} > 10^{-5}$ or (*right*) $m > 10^8\,h^{-1}M_\odot$ at $z = 0$. The meanings of the curves and shaded regions are consistent with those of Fig. 4.2. The disc dramatically reduces the mean abundance of galaxy-hosting subhalo candidates in the halo interior. Nonetheless, the mean curves of all disc configurations lie within the halo-to-halo variance of the disc-less results.

show the distribution of changes in subhalo mass relative to the no-disc (nd) configuration, expressed via the ratio $m_{\rm d}/m_{\rm nd}$, as a function of the most recent orbital pericentric radius, $r_{\rm p}$. Each panel corresponds to a different final disc mass, as indicated. We also plot the median $m_{\rm d}/m_{\rm nd}$ in each $r_{\rm p}$ bin in order to better highlight the trend. The $r_{\rm p}$ are measured directly from the subhalo position data stored in the `SatGen` snapshots. For the purpose of this plot, we measure the $r_{\rm p}$ of each subhalo from the disc-less configuration. However, we acknowledge that the $r_{\rm p}$ of each subhalo is slightly reduced in the presence of a disc,[3] an effect which itself drives a minor enhancement in mass loss. The population of subhaloes included in this analysis have $V_{\rm max,acc} > 30$ km/s, are first-order at $z = 0$, have $r < r_{\rm vir,0}$ and $m/m_{\rm acc} > 10^{-5}$ at $z = 0$ in both the halo-only and halo–disc configuration, and must have experienced at least one pericentric passage. The figure clearly demonstrates that subhaloes that pass closer to the halo centre experience greater mass loss due to the enhanced central density of the halo–disc system than those that are confined to the halo outskirts. While the median $m_{\rm d}/m_{\rm nd}$ begins to deviate from unity for $r_{\rm p} \lesssim 50 - 75$ kpc in all cases, there is a minor indication that the influence of the disc extends out to further radii as its mass is increased. At $r_{\rm p} \approx 15$ kpc, a disc with $f_M = 0.025$ (0.1) drives an additional ~25% (70%) loss of subhalo mass on the median. These results are consistent with a similar analysis by Jiang et al. (2021), although we emphasize that we have used a rather different subhalo selection function in this work. Note that the introduction of the disc slightly changes the orbital period, and hence the orbital phase at $z = 0$, of all subhaloes — this effect results

3. For example, the median (top 1%) reduction in $r_{\rm p}$ due to the $f_M = 0.1$ disc is 0.004 dex (0.045 dex).

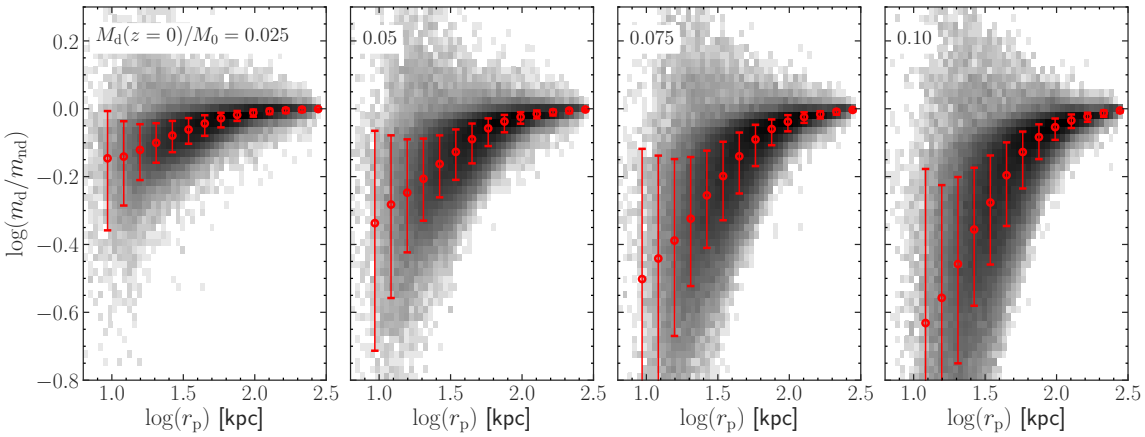

Figure 4.4: The pericentric radius-dependent impact of a disc on the $z = 0$ mass of individual subhaloes. Each panel denotes a different final disc mass, with the remaining disc parameters held to their fiducial values. The most recent pericentric radius, r_p, is measured directly from the `SatGen` outputs of the halo-only configuration. First-order subhaloes with $V_{max,acc} > 30$ km/s, $r < r_{vir,0}$, and $m/m_{acc} > 10^{-5}$ in both the halo-only and halo–disc configuration that have experienced at least one pericentric passage are included. Here, m_d/m_{nd} denotes the ratio between the mass of a particular subhalo in the halo–disc and no-disc configuration. The density heatmap colour is presented on a logarithmic scale. The red circles denote the median m_d/m_{nd} in each r_p bin, whereas the error bars correspond to the 68% interval of halo-to-halo variance. The disc enhances the mass loss of subhaloes that pass near the halo centre while having little impact on those with large-r_p orbits.

in a small fraction of cases where the subhalo mass is actually larger than its counterpart in the disc-less realization at $z = 0$.

In simulation-based subhalo studies, the particle resolution imposes a fixed lower mass limit on the subhalo population. An unfortunate limitation of this approach is that a subhalo is typically inferred to have been "disrupted" after its mass falls below this limit. We have argued in previous studies (van den Bosch & Ogiya, 2018; Green et al., 2021a) that much of this disruption is not physical and is instead a consequence of the simulation mass limit and artificially enhanced by runaway numerical instabilities. Hence, it is instructive to impose a fixed mass limit on the `SatGen` results in order to study the properties of subhaloes whose status as "disrupted" can be specifically traced to the presence of a disc. Here, we define the set of no-disc "survivors" to be all first-order subhaloes with $m/M_0 > 10^{-4}$ and $r < r_{vir,0}$ at $z = 0$ in the no-disc realization. The "disrupted" group is the subset of the "survivors" that instead have $m/M_0 < 10^{-4}$ at $z = 0$ in the halo–disc realization with $f_M = 0.1$ (i.e., the disruption of these subhaloes can be attributed to the presence of the disc). In Fig. 4.5, we present the normalized distributions of several orbital and accretion properties of subhaloes in these two groups. Subhaloes that are most vulnerable to additional disc-driven mass loss, and hence would be preferentially "disrupted" by the disc in a simulation, are simply those on more radial orbits (i.e., smaller circularity, η, as defined in Wetzel, 2011) that pass closer to the halo centre

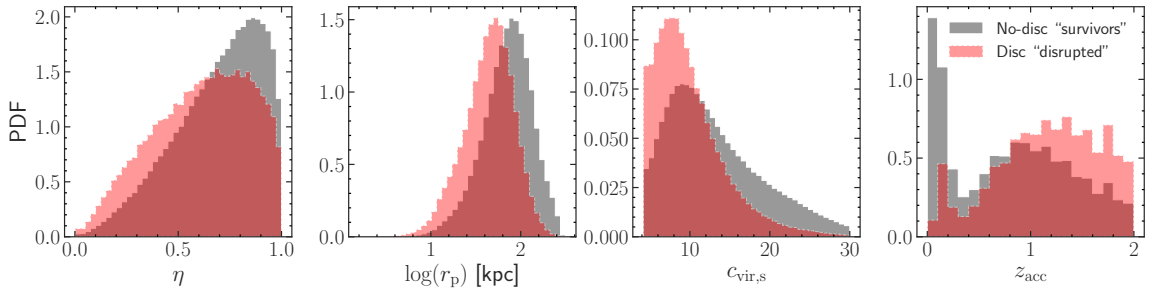

Figure 4.5: From left to right, the normalized distributions of the instantaneous orbital circularity, η, radius of first pericentric passage, r_p, subhalo concentration at accretion, $c_{vir,s}$, and redshift of accretion, z_{acc}, for two groups of subhaloes in the no-disc configuration. The "survivors" group consists of all first-order subhaloes with $m/M_0 > 10^{-4}$ and $r < r_{vir,0}$ at $z = 0$ in the no-disc configuration. The "disrupted" group is the subset of the no-disc "survivors" that have "disrupted" (i.e., $m/M_0 < 10^{-4}$ at $z = 0$) in the composite halo–disc configuration with $f_M = 0.1$. The subhaloes that do not survive the additional disc-driven mass loss tend to be on more radial orbits that penetrate more closely into the host centre (smaller mean η and r_p), are less centrally concentrated, and are typically accreted earlier than the no-disc "survivors".

(smaller r_p), are less centrally concentrated at accretion (smaller $c_{vir,s}$), and have undergone tidal evolution for a longer period of time (larger z_{acc}). We expand on the implications of these relatively intuitive findings in the discussion of "disc shocking" (Section 4.5). Note that a consequence of the preferential "disruption" of radially orbiting subhaloes due to the introduction of the disc is a minor change in the subhalo velocity anisotropy towards more circular orbits. The velocity anisotropy profiles predicted by `SatGen` will be the focus of an upcoming follow-up study.

4.4.4 Azimuthal bias of subhaloes

It is well known that satellite galaxies are preferentially distributed along the major axis of their central host galaxy (e.g., Brainerd, 2005; Yang et al., 2006; Azzaro et al., 2007). This "azimuthal bias" is typically interpreted as implying that central galaxies are aligned with their host haloes. In particular, numerous studies have pointed out that subhaloes in DM simulations are preferentially distributed along the major axis of their host halo (e.g., Knebe et al., 2004; Libeskind et al., 2005; Zentner et al., 2005b). Although a small part of this alignment can be attributed to the preferred direction of subhalo accretion along large-scale filaments (e.g., Aubert et al., 2004; Faltenbacher et al., 2008; Morinaga & Ishiyama, 2020), it is mainly due to the fact that host haloes themselves are not spherical (e.g., Wang et al., 2005; Agustsson & Brainerd, 2006; Wang et al., 2008). Hence, as long as central galaxies are roughly aligned with their host haloes, this non-spherical distribution of subhaloes naturally explains the azimuthal bias in the observed distribution of satellite galaxies (e.g., Agustsson & Brainerd, 2006; Kang et al., 2007).

However, since the central galaxy influences the tidal evolution of the subhaloes, an alternative expla-

nation for the azimuthal bias may be that the central (disc) galaxy preferentially destroys satellite galaxies along more polar-inclined orbits. This would introduce deviations from azimuthal symmetry even when the host halo is spherical and subhaloes are accreted isotropically (as in `SatGen`).

In order to explore this, we search for an angular bias in our predicted subhalo population as a function of disc mass. In `SatGen`, the cylindrical z-axis is the disc axis of symmetry; hence, a disc-driven bias should manifest itself in the form of an asymmetry in the number of polar subhaloes, with $|\cos(\theta)| > 0.5$, and planar subhaloes, with $|\cos(\theta)| < 0.5$, where θ is the host-centric spherical polar angle. We restrict our sample to first-order subhaloes with $m/M_0 > 10^{-4}$ at $z = 0$ that lie within 50 kpc of the host centre, but we emphasize that we considered a range of subhalo selection functions and bounding radii and found qualitatively identical results. In order to compute a robust estimate of the mean polar subhalo fraction, $\langle N(|\cos(\theta)| > 0.5)/N \rangle$, and its uncertainty, we stack our subhalo sample over the ensemble of hosts and perform bootstrap resampling. For each disc mass, we generate 2,000 bootstrap estimates of $N(|\cos(\theta)| > 0.5)/N$ and present the 2.5–97.5 percentile intervals in Fig. 4.6. Note the complete lack of any significant azimuthal bias; on average, there are equal numbers of 'polar' and 'planar' subhaloes.

We ascribe this lack of azimuthal bias to two effects. First of all, the mass loss of subhaloes depends only weakly on the (polar) angle of incidence between the orientation of the disc and the subhalo orbit. This is demonstrated explicitly in Fig. 4.7, which shows the $m(t)/m_{\rm acc}$ trajectories of subhaloes in idealized N-body simulations (see Section 4.3 for details). Different curves correspond to different orbital inclinations, $i = 90^\circ - \theta$, of the initial orbital plane, as indicated, with all other parameters kept fixed. Note that the chosen orbit is highly eccentric, with a small pericentric radius of $r_{\rm p} = 25$ kpc and orbital circularity of $\eta = 0.244$. For comparison, the blue curve shows the corresponding result in the absence of a central disc and the green curve corresponds to the case where the disc has been replaced by a Plummer (1911) sphere with a nearly equivalent spherically enclosed mass profile, which we discuss in Section 4.5.[4] Note that more planar orbits (i.e., those with smaller i) result in slightly *more* mass loss. Hence, if anything, disc-driven disruption should result in a deficit of planar satellites relative to polar satellites, opposite to the trend seen in observational data. The fact that no azimuthal bias emerges is owed to the fact that in an axisymmetric potential the subhalo is not confined to an orbital plane; unlike an orbit in a spherical potential, its polar angle evolves with time. This washes out the weak dependence on the (initial) inclination seen in Fig. 4.7.

To summarize, we conclude that the angular bias observed in the azimuthal distribution of satellite galaxies does not have its origin in a disc-driven preferential disruption of subhaloes along more polar-

4. The Plummer scale length that yields the best match to the spherically enclosed mass profile of the MN disc (when $b_{\rm d}/a_{\rm d} = 0.06$) is $\epsilon \approx 0.92 a_{\rm d}$. The total mass of the two systems is identical but the Plummer (1911) sphere is slightly more centrally concentrated — its enclosed mass is ∼10% larger than that of the disc at $r \approx 10 a_{\rm d}$.

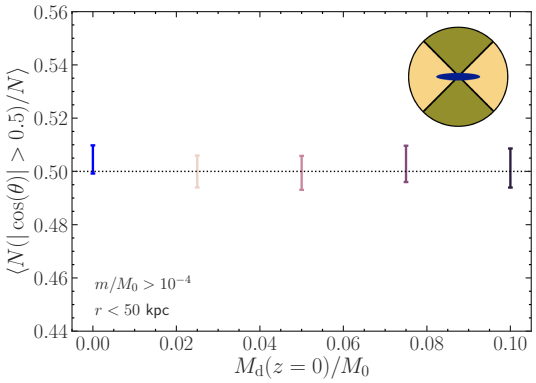

Figure 4.6: The polar subhalo fraction as a function of the disc mass fraction. The sample includes first-order subhaloes with $m/M_0 > 10^{-4}$ at $z = 0$ that lie within 50 kpc of the host centre and is stacked over the ensemble of hosts. The estimated uncertainty range of the fraction of polar subhaloes, which have host-centric polar angles that satisfy $|\cos(\theta)| > 0.5$ (i.e., those in the green region of the schematic in the upper right), is computed via bootstrap resampling. The error bars denote the 2.5–97.5 percentile intervals of the 2,000 bootstrap estimates of the polar fraction. We find no statistically significant disc-driven azimuthal bias in subhalo positions, regardless of disc mass.

inclined orbits. Rather, it is simply due to the central galaxy being aligned with the moment of inertia of the non-spherical host halo combined with the existence of a preferred direction of subhalo accretion due to large-scale filaments.

4.4.5 Dependence on disc parameters

In all previous results, we have studied disc configurations with various final masses (controlled by f_M) but with only the fiducial f_a, β_a, β_M, and $b_{\rm d}/a_{\rm d}$. In Fig. 4.8, we use a summary statistic to demonstrate that our results are insensitive to these other parameters, which control the disc growth and size. For each disc configuration, we compute the mean number of subhaloes (with $V_{\rm max,acc} > 30$ km/s and $m/m_{\rm acc} > 10^{-5}$) enclosed within 50 kpc of the halo centre at $z = 0$ (averaged over all 10,000 trees), which we denote $\langle N(r < 50\,{\rm kpc})\rangle$. In panels 1, 2, 3, 5, and 6, we vary one of the disc parameters, fixing the other four to a set of baseline values ($f_a = 0.0125$, $\beta_a = 1/3$, $f_M = 0.1$, $\beta_M = 1$, and $b_{\rm d}/a_{\rm d} = 0.08$). Note that we use the large $f_M = 0.1$ for our baseline in order to enhance the sensitivity of our results to the other disc parameters. We explore the impact of adiabatic contraction of the host halo in the remaining three panels, which we discuss in Section 4.4.6. For comparison, the horizontal lines indicate the $\langle N(r < 50\,{\rm kpc})\rangle$ of the no-disc configuration, while the gray shaded regions mark the corresponding 68% and 95% halo-to-halo variance intervals. Once again, the mean result lies within the disc-less halo-to-halo variance for every disc configuration studied. Clearly, the mean subhalo abundance within 50 kpc is minimally impacted by the disc scale height and the rate at which the disc grows (both in physical size and mass) relative to the host halo. It

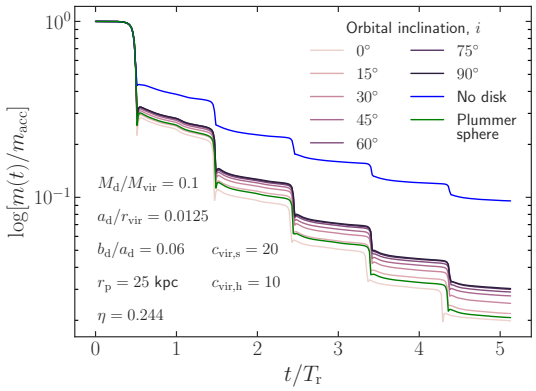

Figure 4.7: The $m(t)/m_\mathrm{acc}$ trajectories of subhaloes in idealized N-body simulations (see Section 4.3). The times are normalized by the radial orbital period, T_r. The disc mass, disc shape, host and subhalo concentrations, orbital energy, and orbital angular momentum are all held fixed as we vary the orbital inclination angle from 0° (the orbit is in the plane of the disc) to 90° (perpendicular to the plane of the disc). As the orbit becomes less inclined, the cumulative mass loss increases, but the overall inclination dependence is weak. Replacing the disc potential with a Plummer (1911) sphere (with the same total mass and a nearly equivalent spherically enclosed mass profile) yields greater mass loss than those of the inclined orbits in the presence of a disc, demonstrating the insignificance of "disc shocking".

is only slightly sensitive to the disc scale length; a more compact disc suppresses more subhaloes. However, the disc mass is the only parameter that has a strong effect on $\langle N(r < 50\,\mathrm{kpc})\rangle$ — the mean abundance drops by 23% (56%) relative to the disc-less case when $f_M = 0.025$ (0.1). For comparison, the replacement of the $f_M = 0.1$ disc with an Einasto (1965) sphere drives a 59% suppression in $\langle N(r < 50\,\mathrm{kpc})\rangle$ relative to the disc-less case. This finding is in excellent agreement with the cosmological simulation study of Garrison-Kimmel et al. (2017), who embed a variety of different disc potentials into their host haloes, finding that only the total mass of the disc has a significant impact on the resulting subhalo statistics.

4.4.6 Adiabatic contraction of the host

Thus far, we have neglected to consider the adiabatic contraction of the host halo due to the formation of the galactic disc. This simplification has enabled us to assess the relative impact of a disc potential on the subhalo population while keeping all other properties of the host consistent with its disc-less counterpart. In addition, proper modeling of adiabatic contraction due to the growth of an axisymmetric potential itself remains an open problem. For example, the 'standard' adiabatic invariant, $rM(r)$, originally suggested by Barnes & White (1984) and Blumenthal et al. (1986) and often used in modeling disc galaxies and their rotation curves (e.g., Mo et al., 1998; van den Bosch & Swaters, 2001), is only valid for unrealistic, completely spherical systems in which all particles move on circular orbits. Furthermore, the fact that the scatter in the Tully–Fisher relation is independent of size (e.g., Courteau & Rix, 1999; Courteau et al., 2007) has been

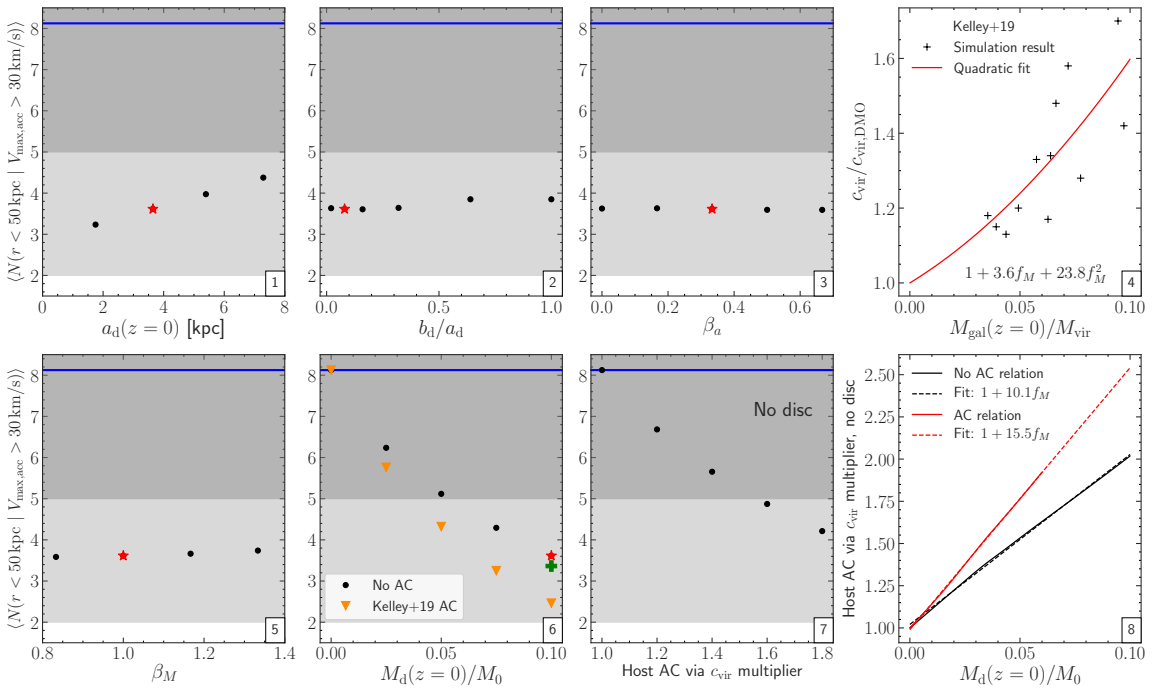

Figure 4.8: The mean abundance of subhaloes with $V_{\mathrm{max,acc}} > 30$ km/s and $m/m_{\mathrm{acc}} > 10^{-5}$ enclosed within 50 kpc of the halo centre at $z = 0$, averaged over all 10,000 merger trees. In panels 1, 2, 3, 5, and 6, we vary one disc parameter, which includes (i) the disc scale length, $a_\mathrm{d}(z = 0)$, which is set by f_a, (ii) the disc scale height, which is expressed as a fraction of the scale length, $b_\mathrm{d}/a_\mathrm{d}$, (iii) the power law slope of $a_\mathrm{d}(z)$, β_a, (iv) the power law slope of $M_\mathrm{d}(z)$, β_M, and (v) the disc mass fraction, $M_\mathrm{d}(z = 0)/M_0$, which is set by f_M. The red star corresponds to the baseline disc configuration ($f_a = 0.0125$, $\beta_a = 1/3$, $f_M = 0.10$, $\beta_M = 1$, and $b_\mathrm{d}/a_\mathrm{d} = 0.08$) around which the parameters are varied. The green cross indicates the spherical Einasto (1965) substitute for the baseline disc. The blue line denotes the $\langle N(r < 50\,\mathrm{kpc})\rangle$ of the halo-only configuration, which is surrounded by its 68% and 95% halo-to-halo variance intervals (dark and light shaded regions, respectively). Consistent with Garrison-Kimmel et al. (2017), only the disc mass has a strong impact on the subhalo statistics. Panel 4 displays the relationship (and corresponding quadratic fit) between the mass fraction of a central galactic potential and the corresponding boost in the NFW concentration of the host due to adiabatic contraction seen in the cosmological simulations of Kelley et al. (2019). This host contraction due to the disc is accounted for in the orange triangles in panel 6. The impact of an increased host concentration in the *absence* of a disc is shown in panel 7. The degeneracy between host contraction (without a disc) and a disc potential (both with and without adiabatic contraction) is demonstrated in panel 8 — increasing the disc mass fraction by 0.01 and ignoring (accounting for) adiabatic contraction due to the disc has the same effect on $\langle N(r < 50\,\mathrm{kpc})\rangle$ as that of a 10% (15.5%) increase in the host concentration.

used to argue that disc formation cannot be associated with significant halo contraction (e.g., Dutton et al., 2007), which might have its origin in non-adiabatic processes operating during disc formation (see e.g., El-Zant et al., 2001b; Tonini et al., 2006).

Despite these issues, we now proceed to investigate how adding (adiabatic) contraction of the host halo due to the assembly of the disc impacts the subhalo population. Rather than assuming a particular adiabatic invariant, we consider a simplified model of the halo contraction based on the simulation results of Kelley et al. (2019). These authors run a suite of twelve dark matter-only cosmological zoom-in simulations, each of which is centred on a different Milky Way-like halo. They then re-run each of the simulations with an embedded galactic potential, which grows over time, placed at the centre of each halo. They fit a NFW profile to each host halo at $z = 0$ in both the halo-only and halo–disc configurations and report the concentrations ($c_{\rm vir,DMO}$ and $c_{\rm vir}$, respectively). In panel 4 of Fig. 4.8, we plot the ratio, $c_{\rm vir}/c_{\rm vir,DMO}$, as a function of the fraction of mass in the embedded potential, $M_{\rm gal}(z = 0)/M_{\rm vir}$. Note that Kelley et al. (2019) model the galaxy as a composite potential that consists of a stellar disc, a gaseous disc, and a stellar bulge; we define $M_{\rm gal}$ to be the combined mass of these three systems and compare it directly with the mass of our single-component stellar disc. Clearly, when the central galaxy makes up a larger fraction of the total host mass, the host experiences greater contraction, which corresponds to a larger $c_{\rm vir}/c_{\rm vir,DMO}$. We fit a quadratic to the relationship, demanding the physical constraint that $c_{\rm vir}/c_{\rm vir,DMO} = 1$ when $M_{\rm gal}(z = 0)/M_{\rm vir} = 0$. We use this fit as the basis of our adiabatic contraction model.

We emulate adiabatic contraction using the following approach. Given the f_M of the disc potential of interest, we look up the corresponding $c_{\rm vir}/c_{\rm vir,DMO}$ using the fit in panel 4 of Fig. 4.8. Our composite host system is exactly the same as before except that the concentration of the NFW host halo is multiplied by the corresponding value of $c_{\rm vir}/c_{\rm vir,DMO}$ at all times during the evolution of the subhaloes. For $f_M = 0.05$ (0.1), the host concentration is boosted by a factor of 1.24 (1.60). Note that the total mass of the composite system enclosed within $r_{\rm vir}$ remains unchanged. In panel 6 of Fig. 4.8, the orange triangles show the mean subhalo abundance within 50 kpc of the host centre, $\langle N(r < 50\,{\rm kpc})\rangle$ (a good summary of the overall influence of the disc), as a function of f_M when the host concentration is boosted in order to account for adiabatic contraction. These can be directly compared to the case without adiabatic contraction for each f_M, denoted by the black circles. As expected, as the disc becomes more massive, the relative impact of the host contraction becomes more significant. Relative to the halo-only host, the $f_M = 0.05$ (0.1) disc suppresses $\langle N(r < 50\,{\rm kpc})\rangle$ by 37% (56%) without adiabatic contraction and by 47% (70%) when the host concentrations are boosted according to the Kelley et al. (2019) model. Note that the $\langle N(r < 50\,{\rm kpc})\rangle$ of the most massive disc remains within the halo-to-halo variance of the halo-only configuration even when adiabatic contraction is taken into account, further demonstrating the importance of such variance in subhalo

statistics.

In Section 4.2, we claimed that the impact of adiabatic contraction is degenerate with simply increasing the disc mass. The reason for this is simple: increasing the disc mass increases the central concentration of the host mass, which is the same result as that of adiabatic contraction. In panel 7 of Fig. 4.8, we consider how $\langle N(r < 50\,\mathrm{kpc})\rangle$ of the halo-only configuration is suppressed if we boost the host concentration by a constant factor at all times during subhalo evolution. In the absence of a disc, boosting the host concentration by a factor of 1.6 drives a 40% reduction in $\langle N(r < 50\,\mathrm{kpc})\rangle$. This level of adiabatic contraction is seen in systems with a disc with $f_M = 0.1$, which itself suppresses $\langle N(r < 50\,\mathrm{kpc})\rangle$ by 56% without the concentration boost. Hence, the effect of the disc itself and the adiabatic contraction that it brings about are both of similar importance, although the disc drives slightly more subhalo suppression. Since the host contraction and disc mass are degenerate, we can use both of panels 6 and 7 in Fig. 4.8 in order to understand the relationship between disc-driven suppression (both with and without also accounting for adiabatic contraction; in terms of $\langle N(r < 50\,\mathrm{kpc})\rangle$) and concentration boost-driven suppression (in the absence of a disc). We interpolate between both the black circles and orange triangles in panel 6 and the black circles in panel 7 and then perform the following: for each value of $M_\mathrm{d}(z = 0)/M_0$ (separately with and without adiabatic contraction), we find the value of the host c_vir multiplier (without a disc) that corresponds to the same $\langle N(r < 50\,\mathrm{kpc})\rangle$. These two relationships are shown in panel 8 of Fig. 4.8 alongside nearly perfect linear fits. In terms of $\langle N(r < 50\,\mathrm{kpc})\rangle$, increasing the disc mass fraction by 0.01 and ignoring (accounting for) adiabatic contraction due to the disc has the same effect as a 10% (15.5%) increase in the host concentration. Hence, the impact of a disc potential can be roughly emulated by simply making the host halo more concentrated.

4.5 Discussion

As we have seen, the presence of a central disc galaxy causes a suppression in the abundance of subhaloes. Our `SatGen`-based results are in good agreement with previous results based on N-body simulations (D'Onghia et al., 2010; Garrison-Kimmel et al., 2017; Errani et al., 2017; Sawala et al., 2017; Kelley et al., 2019), both qualitatively and quantitatively. However, we disagree with D'Onghia et al. (2010) and Garrison-Kimmel et al. (2017) with regards to the importance and origin of this disc-induced substructure depletion. In particular, contrary to these previous studies, we argue that the presence of a disc does not cause actual, physical disruption of substructure. Rather, it merely causes enhanced stripping. We emphasize that this is not just a semantic issue; rather, it is the difference between having no substructure within the inner 30 kpc of the Milky Way and having thousands of low mass subhaloes (a small fraction of which may host a satellite

galaxy) with reduced mass compared to a case without a disc.

In particular, D'Onghia et al. (2010) argued that subhaloes that pass near a central disc galaxy will be destroyed due to impulsive "disc shocking". Whenever a subhalo passes through the plane of a disc at sufficiently high speed, its internal energy will increase by an amount ΔE that can be calculated analytically (Ostriker et al., 1972; Binney & Tremaine, 2008). The crux of the argument made by D'Onghia et al. (2010) is that whenever this ΔE exceeds the binding energy of the subhalo, $|E_{\rm b}|$, the subhalo is 'certain' to be disrupted. They proceed to show that this condition is met by a significant fraction (~15%) of subhaloes in a realistic, MW-like system.

However, as shown in van den Bosch et al. (2018), it is incorrect to assume that $\Delta E > |E_{\rm b}|$ will result in disruption. What matters is not only the total energy injected but also how that energy is distributed over the constituent particles. Since $\Delta E \propto l^2$, the particles in the outskirts of the subhalo, which need little energy to escape, receive the bulk of the energy injection whereas the particles near the subhalo centre, which need a large amount of energy to escape, receive virtually no energy. As a consequence, subhaloes can actually experience a tidal shock that exceeds many multiples of their binding energy and still survive. In fact, van den Bosch et al. (2018) show that subhaloes with a NFW profile that experience a tidal shock with $\Delta E/|E_{\rm b}| = 1$ (10) only lose roughly 20% (55%) of their mass.

Another argument against disc shocking is the fact that in most cases it is subdominant to 'halo shocking', which is tidal heating due to a high speed pericentric passage with respect to the host halo itself. Indeed, van den Bosch et al. (2018) showed that the average $\Delta E/|E_{\rm b}|$ of a subhalo due to its first pericentric passage in a (disc-less) host halo is about ~1.9, which is larger than the average $\Delta E/|E_{\rm b}|$ due to disc shocking. This is consistent with D'Onghia et al. (2010), who showed that halo shocking dominates over disc shocking in a typical MW-like system except for subhaloes with pericentric radii smaller than ~10 kpc.

The insignificance of disc shocking is also evident from Fig. 4.7. If disc shocking were indeed the dominant factor in disc-driven subhalo depletion, then a subhalo on a highly inclined orbit (relative to the disc) should experience more mass loss than a subhalo whose orbit is confined to the plane of the disc. However, Fig. 4.7 demonstrates that the exact opposite trend is seen in our idealized N-body simulations (see Section 4.3), which is nicely reproduced by our tidal stripping model that is implemented in `SatGen`. Furthermore, replacing the disc potential with an equivalent Plummer (1911) sphere in an idealized simulation (see Fig. 4.7) results in greater mass loss than those of the inclined orbits in the presence of a disc. This is consistent with our `SatGen` results, where Figs. 4.2, 4.3, and 4.8 demonstrate that subhalo statistics are generally insensitive to the replacement of the disc with a spherical system that has an equivalent spherically enclosed mass profile.

In fact, our tidal stripping-based mass-loss model successfully reproduces the subhalo mass evolution

of the idealized halo-only *DASH* simulations, our new composite halo–disc simulations (Section 4.3), and a wide range of results from the *Bolshoi* cosmological simulations (Green et al., 2021a) *without an explicit prescription for tidal shocking*. The model relies upon the tight empirical relationship between the stripped subhalo density profile and the bound mass fraction, which is independent of the details of how the mass was lost (e.g., Hayashi et al., 2003; Peñarrubia et al., 2008; Green & van den Bosch, 2019). The accuracy of our mass-loss model and the strength of the density profile–mass fraction relationship would be substantially reduced if tidal shocking dominated for some orbits (i.e., small $r_{\rm p}$) and tidal stripping dominated for others (i.e., large $r_{\rm p}$).

Based on all of these considerations, we conclude that disc shocking plays, at most, a minor role in the substructure suppression caused by a disc potential. Instead, the presence of a disc greatly increases the central mass concentration of the host, which results in an overall increase in tidal stripping that becomes increasingly significant as $r_{\rm p}$ becomes smaller. This net increase in subhalo mass loss effectively shifts the mean SHMF to the left (see Fig. 4.2). The total subhalo abundance above a particular simulation mass limit is decreased. However, considering the fact that a NFW subhalo should never *fully* disrupt (see e.g., van den Bosch et al., 2018; Errani & Peñarrubia, 2020; Errani & Navarro, 2021), we emphasize that this reduced abundance is not due to "disruption", but is instead a consequence of enhanced mass loss combined with a fixed resolution limit.

4.6 Summary

The demographics of DM substructure depend on both the particle nature of DM and the gravitational interaction between DM and baryons. Hence, in order to understand the dependence on the former, we must be able to properly account for the latter. Much progress has been made towards correctly capturing the manner in which baryons shape the overall DM distribution (e.g., D'Onghia et al., 2010; Zolotov et al., 2012; Brooks et al., 2013; Garrison-Kimmel et al., 2017; Sawala et al., 2017; Kelley et al., 2019). However, a common limitation of these studies is that they are all based on expensive cosmological simulations, which has limited their ability to consider statistically complete halo samples and properly contextualize results in terms of the corresponding halo-to-halo variance. The primary finding of these works is clear: the presence of a galactic disc suppresses subhalo abundance, an effect that becomes stronger towards the halo centre. Since the suppression also increases with increasing disc mass, properly accounting for this disc-driven subhalo depletion is especially important for Milky Way-mass systems, which sit at the peak of the stellar mass–halo mass relation.

In this chapter, we used the `SatGen` semi-analytical modeling framework to assess the impact of a galac-

tic disc potential on the DM subhalo demographics of MW-like hosts. This method is not impacted by issues related to numerical disruption, which still hamper the results of N-body simulations, and allows for the construction of large halo samples, which enables unprecedented statistical power. Using an ensemble of 10,000 merger trees with $M_0 = 10^{12}\,h^{-1}M_\odot$, we generated an equally large sample of evolved subhalo populations using a range of different composite halo–disc potentials. This approach allowed us to isolate the differential influence of the disc by controlling for assembly history variance. Leveraging the computational efficiency of `SatGen`, we explored a wide range of disc parameter space, spanning the disc mass, size, and formation history. We used the resulting subhalo catalogs to study subhalo mass functions and radial abundance profiles. We also measured the relative impact of the disc on the $z = 0$ mass of individual subhaloes as well as examined whether disc-driven subhalo depletion gives rise to an azimuthal bias in the spatial distribution of the subhalo population. Our most notable findings are summarized as follows:

- For a disc mass fraction of $f_M = 0.05$, which is a typical value for a Milky Way-size halo, the normalization of the mean SHMF (of subhaloes with $r < r_{\rm vir,0}$) is suppressed by $\lesssim 10\%$ relative to the no-disc case. When only considering subhaloes within 50 kpc of the halo centre, the mean SHMF normalization is decreased by ∼30%. The level of substructure suppression increases with disc mass. However, the mean disc-driven impact on the SHMF is dwarfed by the halo-to-halo variance in all cases.

- The disc has a considerably larger influence on the subhalo abundance near the halo centre, as evidenced by the mean radial subhalo abundance profiles. For example, the mean abundance of potential galaxy-hosting subhaloes (with $V_{\rm max,acc} > 30$ km/s) is suppressed by ∼40% within 50 kpc of the halo centre relative to the no-disc case when $f_M = 0.05$ but is reduced by ∼60% within 20 kpc. The mean effect of the disc on the radial profile is again eclipsed by the halo-to-halo variance.

- By tracking individual subhaloes across different host halo–disc configurations, we have shown that the presence of a central disc causes excess subhalo mass loss, the strength of which increases with decreasing pericentric radius. For example, at $r_{\rm p} \approx 50$ kpc (20 kpc), a disc with $f_M = 0.05$ drives an additional ∼15% (40%) loss of subhalo mass on the median.

- By imposing a fixed mass resolution limit ($m/M_0 > 10^{-4}$), consistent with simulation-based subhalo studies, we analyzed the orbital and accretion properties of subhaloes that survive until $z = 0$ in the absence of a disc but are "disrupted" (i.e., their m falls below the mass cut) by $z = 0$ in the composite host halo–disc case. On average, these disc-disrupted subhaloes are found to have smaller $r_{\rm p}$ and $c_{\rm vir,s}$ than the overall sample.

- The presence of the disc does not cause an azimuthal bias in the spatial distribution of subhaloes for

any of the disc masses considered ($f_M \leq 0.1$). Therefore, the observed alignment of satellite galaxies with the orientation of their central host is not driven by the presence of a disc, but is instead an outcome of galaxy-halo alignment in non-spherical haloes.

- The overall amplitude of disc-driven subhalo depletion is relatively insensitive to the size of the disc and its detailed formation history (both in terms of its size and mass). Rather, it depends almost exclusively on the final mass of the disc. The replacement of the disc with a spherical system, which has a nearly equivalent spherically enclosed mass profile, of the same total mass yields subhalo statistics that are in excellent agreement with the analogous halo–disc configuration.

- We demonstrated that the impact of a disc potential can be emulated by simply increasing the concentration of the host halo. Increasing the disc mass fraction by 0.01 and ignoring (accounting for) adiabatic contraction due to the disc has the same impact on $\langle N(r < 50\,\mathrm{kpc})\rangle$ as boosting the host concentration by 10% (15.5%). Adiabatic contraction of the host due to the formation of the galactic disc only has a significant effect on the overall subhalo abundance when the disc mass fraction is large.

Overall, our `SatGen`-based results are in excellent agreement (both qualitatively and quantitatively) with previous results based on N-body simulations (D'Onghia et al., 2010; Garrison-Kimmel et al., 2017; Errani et al., 2017; Sawala et al., 2017; Kelley et al., 2019). However, as discussed in detail in Section 4.5, we disagree with the notion promoted by D'Onghia et al. (2010) and Garrison-Kimmel et al. (2017) that the disc causes actual disruption of subhaloes via impulsive disc shocking. Rather, the disc simply increases the density in the central region of the halo, which promotes excess mass loss. When this enhanced mass loss results in the subhalo mass dropping below the resolution limit of a numerical simulation, the subhalo appears to have been disrupted; in reality, it would continue to survive with a reduced mass (and should never *fully* disrupt; see e.g., van den Bosch et al., 2018; Errani & Peñarrubia, 2020; Errani & Navarro, 2021). Another new insight that has emerged from this study relates to the overall importance of disc-driven subhalo depletion. By using a large ensemble of merger trees, we were able to demonstrate that the impact of the disc is small compared to the expected halo-to-halo variance, even for the most massive discs considered. Hence, when using the abundance of satellite galaxies or subhaloes in a single system, such as the Milky Way, it is more important to account for halo-to-halo variance than the impact of the central galaxy when making inferences.

Chapter 5

Sunyaev-Zel'dovich effect

This chapter has been published as an article by Green et al. (2020) in the Monthly Notices of the Royal Astronomical Society by Oxford University Press.

5.1 Background

ALAXY clusters are the largest gravitationally bound objects in the Universe, forming hierarchically through accretion at the intersection of cosmic filaments. Their mass- and redshift-distribution is intimately connected to the underlying cosmological model. Hence, a precise approach to linking cluster observables (such as X-ray luminosity or the Sunyaev & Zeldovich (SZ; 1972) effect in the microwave) to the underlying halo mass is essential for using cluster counts as a cosmological probe (Allen et al., 2011; Pratt et al., 2019, for a recent review).

In the upcoming years, the observed X-ray and SZ cluster samples are forecast to grow tremendously. In the X-ray, the recently-launched *eROSITA* mission is set to discover $\gtrsim 10^6$ groups and clusters (Pillepich et al., 2018). In the microwave, the Simons Observatory — planned to begin observations in the early 2020s — will detect the SZ signal of $\gtrsim 10^5$ clusters out to high redshifts (Ade et al., 2019), a catalog that will eventually be augmented to $\gtrsim 10^6$ objects by the next-generation CMB-S4 project (Abazajian et al., 2019). The statistical precision of these surveys will enable unprecedentedly tight cosmological constraints, further stress-testing the standard model of cosmology and potentially illuminating the signal of massive neutrinos or dynamical dark energy. Unlocking the full statistical potential of these surveys necessitates the mitigation of systematic uncertainties associated with cluster gas physics, motivating the development of new halo mass proxies with reduced intrinsic scatter and bias relative to current techniques.

The X-ray luminosity, L_X, is a direct, low-cost mass estimator, but it suffers from high intrinsic scatter

due to poorly-understood cluster core physics. This scatter can be reduced via core-excision (Maughan, 2007; Mantz et al., 2018) or modeling (Käfer et al., 2020) at the cost of sacrificing a significant fraction of the total X-ray photon distribution that comes from the cluster core regions. The integrated SZ signal, Y_{SZ}, is predicted to have a low intrinsic scatter (10 – 15% at fixed mass, e.g., Nagai, 2006; Battaglia et al., 2012) and is much less sensitive to cluster core physics, as the SZ signal arises from the gas permeating throughout the virialized region of galaxy clusters. The product of the X-ray core-excised spectral temperature and gas mass, Y_X (Kravtsov et al., 2006), has comparable scatter to Y_{SZ}, but is only obtainable with high-resolution, long-exposure observations of massive, nearby clusters.

Both X-ray and SZ mass proxies are also subject to scatter due to inter-cluster variance in halo mass accretion histories (MAHs; e.g., Hoekstra et al., 2012; Krause et al., 2012; Barnes et al., 2017b), which results in the presence of varying levels of non-thermal pressure support (Lau et al., 2009; Nelson et al., 2014b). However, the X-ray signal is further afflicted by cooling and heating mechanisms (Stanek et al., 2010), gas clumping (e.g., Nagai & Lau, 2011; Zhuravleva et al., 2013; Khedekar et al., 2013), temperature inhomogeneities (e.g., Rasia et al., 2014), and the cluster dynamical state (e.g., Ventimiglia et al., 2008), whereas the SZ signal is expected to be less sensitive to these details (e.g., Motl et al., 2005; Wik et al., 2008; Eckert et al., 2015, but see also Marrone et al. 2012). Recent machine learning-based efforts have illustrated that the scatter can be reduced by accounting for the dynamical state via full X-ray images (Ntampaka et al., 2019a) or summary statistics of the cluster morphology (Green et al., 2019). Understanding the covariance among these multiple observables will be important for constraining cosmological parameters using multi-wavelength cluster surveys (Stanek et al., 2010).

In addition to introducing scatter, non-thermal pressure support is responsible for a substantial bias that adversely impacts X-ray- and SZ-based mass proxies. These masses are typically estimated under the assumption of hydrostatic equilibrium (HSE) between the gravitational potential and the observed thermal pressure, which is used in lieu of the total pressure. Because of this assumption, the presence of non-thermal pressure in the cluster introduces a HSE mass bias, resulting in observed X-ray/SZ-based masses that are up to 30% lower than the corresponding gravitational lensing-based masses (e.g., Zhang et al., 2010; Mahdavi et al., 2013; von der Linden et al., 2014b; Applegate et al., 2014; Hoekstra et al., 2015; Medezinski et al., 2018; Miyatake et al., 2019). Recent observational studies, however, have shown that this bias is much lower for relaxed populations of clusters that have not recently experienced a significant merger event (e.g., Applegate et al., 2016; Eckert et al., 2019; Ettori et al., 2019; Ghirardini et al., 2019).

To date, the HSE mass biases of simulated clusters have been estimated to be 5 – 40% (e.g., Evrard, 1990; Rasia et al., 2006; Nagai et al., 2007; Lau et al., 2009; Battaglia et al., 2012; Lau et al., 2013; Nelson et al., 2014a; Shi et al., 2016; Biffi et al., 2016; Henson et al., 2017; Ansarifard et al., 2020; Barnes et al., 2021), revealing that

bulk and turbulent intracluster gas motions driven by halo mergers and accretion are likely the dominant source of non-thermal pressure (e.g., Nelson et al., 2012; Avestruz et al., 2016; Shi et al., 2018, 2020) (the potential implications of other sources are discussed in Section 5.4). Measurements of optical weak lensing masses via background galaxies (e.g., Dietrich et al., 2019) and CMB lensing masses (e.g., Raghunathan et al., 2019) may provide a method of calibrating the cluster mass scales and mitigating the HSE bias problem.

As we approach the low-noise, high-resolution frontier of CMB survey science (Mroczkowski et al., 2019), the SZ effect offers promising potential as a cosmological probe. In contrast to X-ray mass proxies, the SZ-based approach suffers from fewer astrophysical systematics and has greater sensitivity to high-redshifts and cluster outskirts. However, as discussed above, assembly-driven non-thermal pressure support is a dominant systematic impeding SZ science. Hydrodynamical simulations demonstrate that the cluster outskirts, which contribute the majority of the thermal SZ (tSZ) signal, have non-thermal pressure support similar in magnitude to the thermal pressure (e.g., Nelson et al., 2014b; Vazza et al., 2018; Walker et al., 2019, for a recent review). In addition to contributing to the scatter and bias in $Y_{\rm SZ}$-based mass estimation, non-thermal pressure also impacts the tSZ angular power spectrum, C_l, which is extremely sensitive to the matter density fluctuation amplitude, $C_l \propto \sigma_8^{7-8}$ (Komatsu & Seljak, 2002). Simulation studies have demonstrated that properly accounting for non-thermal pressure can change the SZ power spectrum amplitude by ~60% (Battaglia et al., 2010; Shaw et al., 2010; Trac et al., 2011), impacting constraints on σ_8 and dark energy (Bolliet et al., 2018). Cross-correlation analyses of SZ, lensing, and galaxy surveys have also been used to constrain the HSE mass bias (e.g., Makiya et al., 2018, 2020; Osato et al., 2020) as well as the roles of AGN feedback and non-thermal pressure of the warm-hot diffuse baryons in groups and clusters (e.g., Van Waerbeke et al., 2014; Battaglia et al., 2015; Hojjati et al., 2017; Osato et al., 2018). Hence, accurately characterizing the average non-thermal pressure profile as a function of cluster mass and redshift is crucial for both subjugating the HSE mass bias problem and using auto- and cross-correlation statistics from upcoming SZ surveys for cosmology. In addition, studying how diversity in halo assembly drives the scatter in the non-thermal pressure support and SZ signal may inform techniques for constructing a more powerful, lower-scatter SZ-based mass proxy that could ultimately strengthen next-generation cosmological analyses.

In this chapter, we study analytically the impact of structure formation-generated turbulence on the scatter in the SZ effect observable-mass scaling relation ($Y_{\rm SZ} - M$). This is made possible by combining the Komatsu & Seljak (2001) model of the cluster total pressure and gas density profiles, the Shi & Komatsu (2014) analytical model of the mass assembly-driven non-thermal pressure profiles, and both average halo MAHs (van den Bosch et al., 2014) and individual Monte Carlo-generated MAHs (Parkinson et al., 2008). Along the way, we identify a near-universality of the average non-thermal pressure fraction profiles, $f_{\rm nth}(r)$,

at fixed peak height that was first hinted at in the simulations of Nelson et al. (2014b). We then calculate the impact of mass assembly on the HSE mass bias, finding that the average bias should increase considerably with both halo mass and redshift due to larger rates of recent mass accretion. Using the thermal pressure profiles computed for various cluster samples, we investigate the slope, normalization, and intrinsic scatter of the $Y_{SZ} - M$ relation and its dependence on aperture radius, redshift, cosmology, and halo mass limit. Importantly, we show that a substantial fraction of the scatter seen in simulated and observed $Y_{SZ} - M$ relations can be attributed to inter-cluster variance in the MAHs. Lastly, we identify a strong correlation between the $Y_{SZ} - M$ residuals and the recent halo mass accretion rate over the previous dynamical time, a relationship that may enable estimation of the accretion rate in observed clusters.

This chapter is organized as follows. In Section 5.2, we describe our methodology, briefly reviewing models of the cluster gas and thermal pressure profiles and the MAHs as well as defining our observables of interest. In Section 5.3, we lay out the results of our analyses, including the predictions for cluster non-thermal pressure profiles (Section 5.3.1), HSE mass biases (Section 5.3.2), observable-mass relationships (Section 5.3.3), and the connection between $Y_{SZ} - M$ residuals and the recent halo mass accretion rate (Section 5.3.4). We discuss the implications of the model in Section 5.4, concluding with a summary of our findings and a forecast of future work in Section 5.5.

The fiducial cosmology used throughout this work is consistent with the Planck Collaboration et al. (2020) results: $\Omega_m = 0.311$, $\Omega_\Lambda = 0.689$, $\Omega_b h^2 = 0.0224$, $h = 0.677$, $\sigma_8 = 0.810$, and $n_s = 0.967$. The base-10 logarithm is denoted by log and the natural logarithm is denoted by ln. Much of the analysis utilizes the COLOSSUS Python package (Diemer, 2018).

5.2 Methods

In this section, we present our analytical framework that we use to model the impact of the assembly history on cluster observables. We first present the theoretical model of the observable-mass scaling relations (Section 5.2.1), which is based on the Kaiser (1986) self-similar model. The cluster observables considered in this study are all functions of the gas density, temperature, and thermal pressure in the intracluster medium (ICM). In Section 5.2.2, we describe the techniques used to generate the MAHs of individual clusters (Cole et al., 2000; Parkinson et al., 2008) and their population averages (van den Bosch et al., 2014), enabling us to study both mean trends and quantify inter-cluster variance. We assume that the gas density and *total* pressure are well-described by the model of Komatsu & Seljak (2001), which we present in Section 5.2.3. The thermal pressure is obtained from the total by subtracting off the non-thermal component, which we compute using the model of Shi & Komatsu (2014), presented in Section 5.2.4. We assume throughout that the

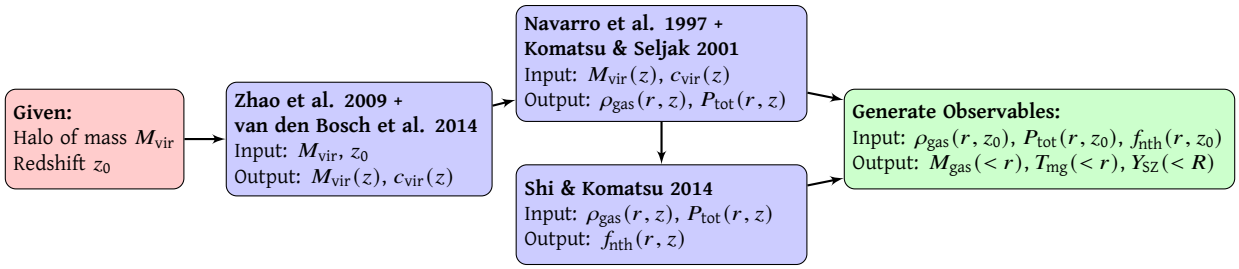

Figure 5.1: A flowchart that summarizes our theoretical framework. For each halo with a virial mass of $M_{\rm vir}$ observed at a redshift of z_0, the mass accretion history and concentration history are generated following Section 5.2.2. This is input into the gas model (Section 5.2.3), which assumes hydrostatic equilibrium, in order to generate the gas density and total pressure profiles throughout the accretion history. The non-thermal pressure fraction profile is then generated following Section 5.2.4. Lastly, the gas density profile and total/non-thermal pressure profiles are used to generate the observables: gas mass, temperature, and integrated SZ signal.

non-thermal pressure is entirely due to turbulence generated during the cluster's mass assembly. Lastly, we lay out our methods used to compute and quantify the properties of cluster scaling relations in Section 5.2.5. The model framework is summarized in Fig. 5.1.

5.2.1 Observables and self-similar scaling relations

Our main goal is to model the scaling relation between the observable, cylindrically-integrated SZ signal, $Y_{\rm SZ}$, and the observationally inferred cluster mass. As discussed below, the SZ signal is proportional to both the cluster gas mass, $M_{\rm gas}$, and the gas mass-weighted temperature, $T_{\rm mg}$. We therefore also analyze the scaling relations between these quantities and cluster mass. We study the dependence of all of these scaling relations on the aperture radius, $r_{\rm ap}$, for which we use multiples of $r_{\rm 500c}$ and $r_{\rm 200m}$.[1] In what follows, we use R to denote two-dimensional projected distances and r to denote three-dimensional distances; in particular, $R_{\rm ap}$ and $r_{\rm ap}$ are used to indicate the aperture radii used for cylindrically- and spherically-integrated quantities, respectively. The total enclosed halo mass is denoted $M(<r_{\rm ap})$. We emphasize that throughout this study we always use the same aperture to compute both the total enclosed halo mass and the cluster observables; however, when studying a cylindrically-integrated observable, we shall still compare it to the spherically enclosed halo mass using the same numerical values for both $r_{\rm ap}$ and $R_{\rm ap}$. The three main observables considered in this chapter are $Y_{\rm SZ}(<R_{\rm ap})$, $M_{\rm gas}(<r_{\rm ap})$, and $T_{\rm mg}(<r_{\rm ap})$.[2] Computing these quantities requires a model of the *thermal* pressure and gas density profiles, $P_{\rm th}(r)$ and $\rho_{\rm gas}(r)$, which we

1. Note that $r_{\rm 500c}$ is the radius inside of which the mean density is equal to 500 times the *critical* density, $\rho_{\rm c}(z)$, whereas within $r_{\rm 200m}$, the mean density is 200 times the *mean matter* density, $\rho_{\rm m}(z)$. For cluster mass scales, the virial radius is $r_{\rm vir} \approx 2r_{\rm 500c} \approx 0.8r_{\rm 200m}$ at $z = 0$.

2. Note that of $M_{\rm gas}$, $T_{\rm mg}$, and $Y_{\rm SZ}$, only $Y_{\rm SZ}$ is computed as a two-dimensional projected quantity in this work — hence, our $T_{\rm mg}$ and $M_{\rm gas}$ are not direct observables, but studying these spherically-integrated quantities is still illuminating with regards to understanding the $Y_{\rm SZ} - M$ relation.

describe in Sections 5.2.3 and 5.2.4, respectively.

The self-similar model developed by Kaiser (1986) is a simple model of cluster evolution based on three assumptions: (i) clusters form from peaks in the initial density field of an Einstein–de Sitter universe with $\Omega_{\rm m} = 1$, (ii) the amplitude of the primordial density fluctuations varies with spatial scale as a power law, and (iii) processes that impact cluster formation do not introduce additional physical scales to the problem (Kravtsov & Borgani, 2012). Under these assumptions, the gravitational collapse of galaxy clusters is *self-similar*. However, various astrophysical processes, such as turbulence, introduce additional physical scales, which result in cluster evolution that deviates from self-similarity. Before we investigate how non-thermal pressure support causes deviations from self-similarity, we briefly describe how the relevant observables (i.e., $M_{\rm gas}$, $T_{\rm mg}$, and $Y_{\rm SZ}$) scale in the Kaiser (1986) model.

Given a particular overdensity definition, cluster mass and radius are interchangeable via

$$M_{\Delta} = (4\pi/3)\Delta\rho_x(z)r_{\Delta}^3, \tag{5.1}$$

where Δ is the overdensity factor. When haloes are defined with respect to a multiple of the critical density, $\rho_x(z) \equiv \rho_{\rm c}(z) \propto E^2(z) = \Omega_{\rm m}(1+z)^3 + \Omega_{\Lambda}$, whereas when they are defined with respect to the mean matter density, $\rho_x(z) \equiv \rho_{\rm m}(z) \propto (1+z)^3$.

The Kaiser (1986) model assumes that the density profile of the gas, $\rho_{\rm gas}(r)$, is self-similar and that its normalization is such that, for fixed $r_{\rm ap}$, the ratio between the enclosed gas mass, given by

$$M_{\rm gas}(< r_{\rm ap}) = 4\pi \int_0^{r_{\rm ap}} \rho_{\rm gas}(r) r^2 {\rm d}r, \tag{5.2}$$

and the enclosed total mass, $M(< r_{\rm ap})$, is independent of halo mass. As we will see, the halo concentration-mass relation introduces an additional mass-dependence that causes the gas profile shapes to deviate from self-similarity. However, for large $r_{\rm ap}$, the assumption of a fixed $M_{\rm gas}(< r_{\rm ap})/M(< r_{\rm ap})$ is still reasonable because recent findings in both simulations and observations have found that the cumulative gas mass fraction approaches the cosmic baryon fraction at or below $\sim(1-2)r_{\rm 200m}$ (e.g., Kravtsov et al., 2005; Ettori et al., 2006; Planelles et al., 2013; Eckert et al., 2013; Mantz et al., 2014; Morandi et al., 2015) for clusters with $M_{\rm 500c} \gtrsim 10^{14}\, h^{-1} M_{\odot}$ at $0 \lesssim z \lesssim 1$. This assumption is less realistic for lower mass haloes, where gas depletion due to feedback becomes significant.

The Kaiser (1986) model assumes that the gas is in HSE with the gravitational potential and that the logarithmic slopes of the gas density and thermal pressure profiles are independent of halo mass. Hence,

from the HSE equation,

$$\frac{1}{\rho_{\rm gas}(r)}\frac{{\rm d}P(r)}{{\rm d}r} = -\frac{{\rm d}\Phi(r)}{{\rm d}r} = \frac{GM(<r)}{r^2}\,, \tag{5.3}$$

where $P(r) = P_{\rm th}(r)$ in the absence of non-thermal pressure, and assuming an ideal gas, we have that

$$M(<r) = -\frac{k_{\rm B}\,T(r)\,r}{\mu m_{\rm p}\,G}\left[\frac{{\rm d}\ln T}{{\rm d}\ln r} + \frac{{\rm d}\ln \rho_{\rm gas}}{{\rm d}\ln r}\right], \tag{5.4}$$

where $\mu m_{\rm p}$ is the mean particle mass. This equation can be used to solve for $T(r)$ given $M(<r)$ and $\rho_{\rm gas}(r)$, from which we compute the spherically-integrated gas mass-weighted temperature

$$T_{\rm mg}(<r_{\rm ap}) = \frac{4\pi}{M_{\rm gas}(<r_{\rm ap})}\int_0^{r_{\rm ap}} \rho_{\rm gas}(r)\,T(r)\,r^2\,{\rm d}r\,. \tag{5.5}$$

For the self-similar gas density profile assumed in the Kaiser (1986) model, this yields the following scaling relation

$$T_{\rm mg}(<r_{\rm ap}) \propto M(<r_{\rm ap})^{2/3}\,[\Delta\rho_x(z)]^{1/3}\,. \tag{5.6}$$

Note the dependence on $\Delta\rho_x(z)$, which introduces a redshift dependence in the normalization of this scaling relation between cluster temperature and mass.

Lastly, to compute the cylindrically-integrated SZ signal, $Y_{SZ}(<R_{\rm ap})$, we first calculate the Compton-y parameter by integrating the thermal pressure of the gas along the line-of-sight using

$$y_{\rm SZ}(R) = 2\int_R^{r_{\rm b}} n_{\rm e}(r)\frac{k_{\rm B}T_{\rm e}(r)}{m_{\rm e}c^2}\,\sigma_{\rm T}\,\frac{r\,{\rm d}r}{\sqrt{r^2-R^2}}\,, \tag{5.7}$$

where $k_{\rm B}$, $m_{\rm e}$, c, and $\sigma_{\rm T}$ are standard constants and $n_{\rm e}(r)$ and $T_{\rm e}(r)$ denote the electron gas number density and temperature profiles. The line-of-sight integration is performed out to $r_{\rm b} \equiv 2r_{\rm 200m}$, which is roughly consistent with the radius of the accretion shock beyond which the pressure profile rapidly drops to the ambient pressure of the intergalactic medium (see e.g., Molnar et al., 2009; Lau et al., 2015). We then integrate $y_{\rm SZ}(R)$ over the aperture using

$$Y_{\rm SZ}(<R_{\rm ap}) = 2\pi\int_0^{R_{\rm ap}} y_{\rm SZ}(R)R\,{\rm d}R. \tag{5.8}$$

As can be seen from equation (5.7), the SZ signal is proportional to the product of the gas density and

Quantity	Mass Slope	Overdensity Slope
$Y_{\rm SZ}(< R_{\rm ap})$	$5/3$	$1/3$
$M_{\rm gas}(< r_{\rm ap})$	1	0
$T_{\rm mg}(< r_{\rm ap})$	$2/3$	$1/3$

Table 5.1: The observable cluster quantities studied in this work alongside their predicted power-law coefficients with respect to halo mass and spherical overdensity according to the self-similar model. The integrated Sunyaev-Zel'dovich signal, $Y_{\rm SZ}$, is calculated within a projected aperture, whereas the gas mass, $M_{\rm gas}$, and gas mass-weighted temperature, $T_{\rm mg}$, are computed within a spherical aperture.

temperature. Hence, the self-similar Kaiser (1986) model predicts that

$$\begin{aligned} Y_{\rm SZ}(< R_{\rm ap}) &\propto M_{\rm gas}(< r_{\rm ap})\, T_{\rm mg}(< r_{\rm ap}) \\ &\propto M(< r_{\rm ap})^{5/3}\, [\Delta\rho_x(z)]^{1/3}\,. \end{aligned} \tag{5.9}$$

In Section 5.3.3, we study the deviations of these observable-mass relations from self-similarity due to the injection of turbulence via mass assembly. The observable quantities and their self-similar scaling relations are summarized in Table 5.1.

5.2.2 Mass accretion histories

We assume that the dark matter distribution of haloes follow the Navarro-Frenk-White (NFW) density profile (Navarro et al., 1997) with enclosed mass

$$M(< r) = M_{\rm vir}\frac{f(c_{\rm vir}r/r_{\rm vir})}{f(c_{\rm vir})}, \tag{5.10}$$

where $M_{\rm vir}$, $r_{\rm vir}$, and $c_{\rm vir}$ are the halo virial mass, radius, and concentration,[3] respectively, and $f(x) = \ln(1+x) - x/(1+x)$.

The mass assembly history, $M_{\rm vir}(z)$, tracks the *main branch* of the halo, which is the branch of the halo merger tree that follows the main progenitor of the main progenitor of the main progenitor and so on (halo merger trees are discussed in detail in Section 2.1 of Jiang & van den Bosch, 2016). We compute *individual* MAHs using the merger tree method described in Parkinson et al. (2008), a Monte Carlo approach based on the extended Press–Schechter (EPS; Bond et al., 1991) formalism, which the method comparison project of Jiang & van den Bosch (2014) finds to perform the best at reproducing merger trees in simulations.

The EPS formalism gives the progenitor mass function (PMF), $n(M_{\rm p}, z_2|M_1, z_1)\,{\rm d}M_{\rm p}$, which specifies the average number of progenitor haloes with a mass of $M_{\rm p} \pm {\rm d}M_{\rm p}/2$ present at z_2 that merge into a descendant

3. $M_{\rm vir}$ is the mass enclosed within $r_{\rm vir}$, inside of which the mean density is equal to $\Delta_{\rm vir}(z)$ times the critical density. At $z = 0$, $\Delta_{\rm vir}(z) \approx 100$, and is otherwise well-described by Bryan & Norman (1998) for general z and cosmology. The concentration is $c_{\rm vir} = r_{\rm vir}/r_{\rm s}$, with $r_{\rm s}$ the NFW scale radius.

halo with a mass of M_1 at $z_1 < z_2$. Given a target halo mass of $M_{\rm vir,0}$ at redshift of observation z_0, one can sample a set of progenitor halo masses from the PMF, $M_{\rm p,1}$, $M_{\rm p,2}$, ..., $M_{\rm p,}N$, that at a previous time of $z_1 = z_0 + \Delta z$ satisfy $\sum_{i=1}^{N} M_{{\rm p},i} = M_{\rm vir,0}$. Beginning at z_0, the merger tree is constructed by walking backwards in time with a temporal resolution of Δz (which need not be constant along the tree), at each point sampling the progenitors of each descendant down to a mass resolution of $M_{\rm res}$. This $M_{\rm res}$ is typically a fixed fraction of the target halo mass, which we denote $\psi_{\rm res} = M_{\rm res}/M_{\rm vir,0}$; throughout this work, we use a mass resolution of $\psi_{\rm res} = 10^{-4}$. The Parkinson et al. (2008) method generates merger trees based on the 'binary method with accretion' of Cole et al. (2000) alongside a PMF modified from EPS theory to reproduce the merger statistics of the *Millennium Simulation* (Springel et al., 2005). For the construction of the tree, we use the timestep schedule motivated in Parkinson et al. (2008), which corresponds to $\Delta z \approx 10^{-3}$. However, as discussed in Section 2.2 of van den Bosch et al. (2014), for the purpose of computational efficiency, we down-sample the temporal resolution of the merger tree outputs to a timestep of $\Delta t = 0.1 t_{\rm ff}(z)$. The free-fall time for a halo with a critical overdensity of 200 at a redshift of z, $t_{\rm ff}(z) \propto (1+z)^{-3/2}$, is on the order of the halo dynamical time. Hence, there is little information added by using a smaller Δt; we have verified that our subsequent results are converged with respect to the merger tree timestep.

These Monte Carlo MAHs are used in our analysis of the observable-mass relations in Sections 5.3.3 and 5.3.4. When we are interested in the *average* properties of a given halo of mass $M_{\rm vir,0}$ at redshift of observation z_0, we use the 'universal model' of the average MAH described in van den Bosch et al. (2014) (see their Section 4.1 and Appendix C). In this case, we also trace the MAH back to the redshift that satisfies $\psi_{\rm res} = 10^{-4} = M(z)/M_{\rm vir,0}$. The average MAHs are used to study the properties of $f_{\rm nth}(r)$ and the HSE bias in Sections 5.3.1 and 5.3.2.

For illustrative purposes, in Fig. 5.2, we show many different Monte Carlo-generated $M_{\rm vir}(z)$ trajectories for haloes with $\log(M_{\rm vir}(z=0)/[h^{-1}M_\odot]) = 14$. In addition, we overplot the average MAH predicted by the van den Bosch et al. (2014) model for a halo of the same mass, demonstrating good agreement.

Concentrations are determined using the model of Zhao et al. (2009) as modified by van den Bosch et al. (2014) to accurately reproduce the concentrations seen in *Bolshoi*. The halo concentrations are given by

$$c_{\rm vir}(M_{\rm vir}, t) = 4.0\left[1 + \left(\frac{t}{3.40 t_{0.04}}\right)^{6.5}\right]^{1/8}. \tag{5.11}$$

At proper time t, the halo has mass $M_{\rm vir}(t)$. The time $t_{0.04}$ is the proper time at which the halo's progenitor has accumulated a mass of $0.04 M_{\rm vir}(t)$, which can be computed directly from the MAH. If $0.04 M_{\rm vir}(t) < \psi_{\rm res} M_{\rm vir,0}$, we set $c_{\rm vir}(t) = 4$, which is the lower bound in the Zhao et al. (2009) model that all haloes tend toward at high z; we have verified that our results are insensitive to this choice. Note that each halo has

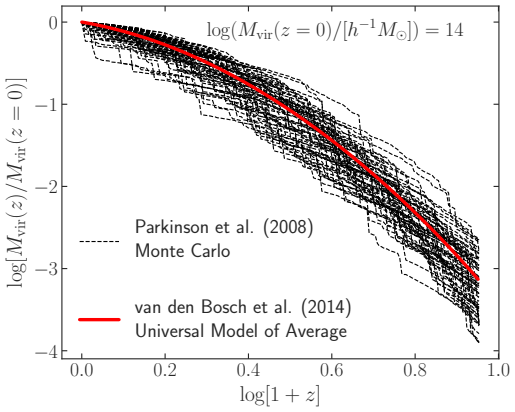

Figure 5.2: Example mass accretion histories, $M_{\rm vir}(z)$, for haloes with a final mass of $\log(M_{\rm vir}(z = 0)/[h^{-1}M_\odot]) = 14$. The black, dashed lines correspond to individual MAHs generated by the Monte Carlo method of Parkinson et al. (2008) whereas the thick, red line represents the 'universal model' of the average MAH developed in van den Bosch et al. (2014). The individual MAHs are essential for studying the scatter in the observable-mass scaling relations (Sections 5.3.3 and 5.3.4), whereas the average MAH is used for studying properties of the non-thermal pressure fractions and HSE mass bias (Sections 5.3.1 and 5.3.2).

a $c_{\rm vir}(t)$ trajectory determined solely by its MAH. We have verified that the main results of this work are insensitive to the specific $c_{\rm vir}(M_{\rm vir}, z)$ model used (we isolate the effect of the $c_{\rm vir}(M_{\rm vir}, z)$ relation on our results in Section 5.3.3).

When discussing the effect of the mass assembly history on deviations from the self-similar observable-mass relations, it is convenient to use a summary statistic of $M_{\rm vir}(z)$ that captures the mass accretion rate (MAR) over a finite period of time. Throughout, we use the definition of the MAR introduced in Diemer (2017), which encapsulates the change in $M_{\rm 200m}$ over one dynamical (or crossing) time, $t_{\rm dyn} = 2r_{\rm 200m}/v_{\rm 200m}$, where $v_{\rm 200m}$ is the circular velocity at $r_{\rm 200m}$. This MAR is written as

$$\Gamma = \frac{\log[M_{\rm 200m}(a_{\rm obs})] - \log[M_{\rm 200m}(a_1)]}{\log(a_{\rm obs}) - \log(a_1)}, \tag{5.12}$$

where $a_{\rm obs} = (1+z)^{-1}$ corresponds to the redshift of observation z and $a_1 = a(t_{\rm obs} - t_{\rm dyn})$ is the scale factor one dynamical time prior to observation. In practice, the MAHs are discretely sampled in time, so we approximate a_1 and $M_{\rm 200m}(a_1)$ as the value of the scale factor and mass at the timestep that is closest to $t(a_1)$ in the MAH output. Note that the halo concentration anti-correlates with Γ — more relaxed systems tend to be more highly concentrated.

5.2.3 Total pressure and gas density profiles

Assuming the dark matter halo is well-described by the NFW density profile, Komatsu & Seljak (2001, hereafter KS01) develop a polytropic gas model for clusters in HSE where the thermal pressure is $P_{\rm th}(r) \propto$

$\rho_{\rm gas}(r)T(r) \propto \rho_{\rm gas}(r)^{\gamma}$, with γ the polytropic index (note that this is different than the adiabatic coefficient).

However, by studying simulated galaxy clusters, Shaw et al. (2010) find that a polytropic model describes the *total* pressure profile, $P_{\rm tot}(r)$, better than $P_{\rm th}(r)$. Specifically, Shaw et al. (2010) report that $P_{\rm tot}(r)$ follows a polytrope with fixed $\gamma = 1.2$ over four decades in $\rho_{\rm gas}(r)$, indicating that γ does not vary with cluster-centric radius. Hence, we use the model of KS01 to compute $P_{\rm tot}(r) \propto \rho_{\rm gas}(r)T_{\rm eff}(r) \propto \rho_{\rm gas}(r)^{\gamma}$. Here, $T_{\rm eff}(r)$ is an effective temperature profile that accounts for both the thermal and non-thermal pressure, which we write as $T_{\rm eff}(r) \equiv T_{\rm eff,0}\theta(r)$. The resulting total pressure and gas density are thus parameterized as

$$P_{\rm tot}(r) = P_0\theta(r)^{\frac{\gamma}{\gamma-1}} \quad \text{and} \quad \rho_{\rm gas}(r) = \rho_0\theta(r)^{\frac{1}{\gamma-1}}, \tag{5.13}$$

where all of P_0, ρ_0, and $\theta(r)$ depend on $M_{\rm vir}$ and $c_{\rm vir}$.[4] In addition, for reasons explained below, we have that $\gamma = \gamma(c_{\rm vir})$. Plugging these parameterizations into the HSE equation (equation [5.3]), where we now use $P(r) = P_{\rm tot}(r)$, yields

$$\theta(r, M_{\rm vir}, c_{\rm vir}) = 1 + \frac{\gamma - 1}{\gamma}\frac{\rho_0}{P_0}\left[\Phi(0) - \Phi(r)\right], \tag{5.14}$$

with $\Phi(r)$ the NFW gravitational potential profile, given by

$$\Phi(r) = -\frac{GM_{\rm vir}}{r_{\rm vir}}\frac{c_{\rm vir}}{f(c_{\rm vir})}\frac{\ln(1 + c_{\rm vir}r/r_{\rm vir})}{c_{\rm vir}r/r_{\rm vir}}. \tag{5.15}$$

A core assumption of KS01 is that the gas density profile traces the dark matter density profile in the outer halo. Under this assumption, the normalization of the mass-temperature relation (or, when $P_{\rm tot} \neq P_{\rm th}$, the mass-effective-temperature relation; i.e., $P_0/\rho_0 \propto T_{\rm eff,0}$) is determined by asserting that the slope of the dark matter and gas density profiles are the same at some matching radius, r_*. In order for the gas profile to trace the dark matter profile over a large radial range ($r_{\rm vir}/2 < r < 2r_{\rm vir}$), their slopes must agree for a range of r_*. Since the value of P_0/ρ_0 should not depend on the choice of r_*, the polytropic index, γ, is set such that P_0/ρ_0 is independent of r_* (in other words, γ solves $\mathrm{d}(P_0/\rho_0)/\mathrm{d}r_* = 0$). Since the shape of the dark matter density profile depends on halo mass via the mass-concentration relation, both γ and P_0/ρ_0 also depend on $c_{\rm vir}$. We follow KS01, adopting their polynomial fitting functions given by their equations (25) and (26). Both γ and P_0/ρ_0 grow with $c_{\rm vir}$ and thus tend to be lower in disturbed systems with high MAR.

Motivated by the discussion of $M_{\rm gas}(< r_{\rm ap})/M(< r_{\rm ap})$ in Section 5.2.1, we set the normalization of

4. The native mass definition of the KS01 gas model and our MAH models (Section 5.2.2) is that of $\Delta_{\rm vir}(z)$. We convert between mass and radius definitions using the concentration model in equation (5.11) and adopt the '200m' and '500c' mass conventions for comparisons with various simulation and observational results.

$\rho_{\rm gas}(r)$ such that $M_{\rm gas}(< 2r_{\rm 200m})$ is equal to the cosmic baryon fraction $\Omega_{\rm b}/\Omega_{\rm m}$ times $M(< 2r_{\rm 200m})$; our results are insensitive to the exact radius used to set this normalization.

5.2.4 Non-thermal pressure profile

In order to calculate cluster observables, we need to disentangle $P_{\rm th}(r)$ from $P_{\rm tot}(r) = P_{\rm th}(r) + P_{\rm nth}(r)$, where $P_{\rm nth}(r)$ is the non-thermal pressure. We determine the non-thermal pressure by following the analytical model of Shi & Komatsu (2014, hereafter SK14). From the KS01 total pressure and gas density profiles, we calculate the total velocity dispersion of the gas (per degree of freedom) as

$$\sigma^2_{\rm tot}(r, M_{\rm vir}, c_{\rm vir}) = P_{\rm tot}(r)/\rho_{\rm gas}(r) = (P_0/\rho_0)\theta(r). \tag{5.16}$$

We emphasize that $\sigma^2_{\rm tot}$ evolves in time due to changing mass *and* concentration, i.e., $M_{\rm vir}(z)$ and $c_{\rm vir}(z)$. The ansatz of SK14 is that the turbulent energy (per unit mass per degree of freedom), $\sigma^2_{\rm nth}(r)$, dissipates on a timescale proportional to the eddy turn-over time of the largest eddies, which is in turn proportional to the local orbital time, $t_{\rm dis}(r) = \beta t_{\rm orb}(r)/2$, and a fraction η of the total energy injected into the cluster via mass growth is converted into turbulence. Based on this ansatz, the non-thermal energy evolves as

$$\frac{{\rm d}\sigma^2_{\rm nth}}{{\rm d}t} = -\frac{\sigma^2_{\rm nth}}{t_{\rm dis}} + \eta\frac{{\rm d}\sigma^2_{\rm tot}}{{\rm d}t}. \tag{5.17}$$

The free parameters are calibrated against cosmological simulations in Shi et al. (2015) to $\beta = 1$ and $\eta = 0.7$, which we adopt throughout. Determining $\sigma^2_{\rm nth}$ at redshift z is an initial value problem; SK14 find that the results are insensitive to the initial redshift, z_i, and initial $\sigma^2_{\rm nth}$, opting to use $z_i = 6$ and $\sigma^2_{\rm nth}(r, z_i) = \eta\sigma^2_{\rm tot}(r, z_i)$. Rather than begin at a fixed z_i, our initial redshift varies based on the z_i that satisfies $M(z_i) = \psi_{\rm res} M_{\rm vir,0}$. For example, for $\psi_{\rm res} = 10^{-4}$, haloes in the mass range $12 \le \log(M_{\rm vir}(z = 0)/[h^{-1}M_\odot]) \le 15.5$ have a z_i that varies from $5 - 20$, with a distribution that peaks at $z_i = 10$. Hence, our initial conditions use $\psi_{\rm res} = 10^{-4}$ and $\sigma^2_{\rm nth}(r, z_i) = \eta\sigma^2_{\rm tot}(r, z_i)$.[5] We have verified that our subsequent results do not change if we decrease $\psi_{\rm res}$ (i.e., increase z_i); additionally, we have verified that at $\psi_{\rm res} = 10^{-4}$, the results are insensitive to the initial $\sigma^2_{\rm nth}(r, z_i)$ profile used. This is because at the corresponding sufficiently high initial redshift, the time between z_i and z is many multiples of the initial $t_{\rm dis}(r)$ (i.e., $15 - 10^4$ times) for all r and z of interest, dissipating away the initial value of $\sigma^2_{\rm nth}$. Note that $\sigma^2_{\rm nth}$ is evolved independently for each cluster-centric radius r.

5. We also impose the physical constraint that whenever $\sigma^2_{\rm nth}(r, t_{j-1}) < -{\rm d}\sigma^2_{\rm nth}(r, t_{j-1})$, then $\sigma^2_{\rm nth}(r, t_j) = 0$ rather than becoming negative; this can happen occasionally near the halo center, where $t_{\rm dis}$ is small.

The timestep used in the $\sigma^2_{\rm nth}$ evolution is the same as that of the merger tree, which corresponds to 10% of the instantaneous cluster free-fall time at a critical overdensity of 200. At each timestep, the halo mass and concentration are updated based on the MAH model described above in Section 5.2.2. The updated $M_{\rm vir}(z_j)$ and $c_{\rm vir}(z_j)$ result in updated $\gamma(c_{\rm vir})$, $(P_0/\rho_0)(c_{\rm vir})$, and $\Phi(r, M_{\rm vir}, c_{\rm vir})$, which we can use to compute $\sigma^2_{\rm tot}(r, z_j)$. We then compute

$$\frac{{\rm d}\sigma^2_{\rm tot}}{{\rm d}t}(r, z_j) = \frac{\sigma^2_{\rm tot}(r, z_j) - \sigma^2_{\rm tot}(r, z_{j-1})}{t(z_j) - t(z_{j-1})}, \tag{5.18}$$

after which we can compute ${\rm d}\sigma^2_{\rm nth}/{\rm d}t$ to get our updated $\sigma^2_{\rm nth}(r, z_j)$. Note that different $M_{\rm vir}(z)$ and $c_{\rm vir}(z)$ trajectories thus result in different $({\rm d}\sigma^2_{\rm tot}/{\rm d}t)(r, z)$ trajectories. Hence, for fixed $M_{\rm vir}$ at observation redshift z, *all variance in the $\sigma^2_{\rm nth}$ profiles is due to inter-cluster differences in MAHs.* We have checked our results for convergence with respect to timestep in the $\sigma^2_{\rm nth}$ evolution, finding that the final $\sigma^2_{\rm nth}$ change insignificantly when the size of the timestep is reduced by a factor of five.

With $\sigma^2_{\rm nth}(r)$ computed, we define the non-thermal energy fraction as $f_{\rm nth}(r) = \sigma^2_{\rm nth}(r)/\sigma^2_{\rm tot}(r)$. The thermal pressure profile is then $P_{\rm th}(r) = [1 - f_{\rm nth}(r)]P_{\rm tot}(r)$.[6] From $P_{\rm th}(r)$ and $\rho_{\rm gas}(r)$, we can compute the aforementioned cluster observables. In addition, we explore the mass and redshift dependence of the non-thermal pressure fraction and its implications for the HSE mass bias in Sections 5.3.1 and 5.3.2, respectively.

We emphasize that all results have been tested for convergence with respect to (i) the temporal resolution of the MAH and associated $\sigma^2_{\rm nth}$ equation integration, (ii) the initial conditions used for the integration of $\sigma^2_{\rm nth}$ (i.e., $\sigma^2_{\rm nth}(r, z_i)$ and $\psi_{\rm res}$), (iii) the spatial resolution of the cluster profiles used to integrate the observables, and, where relevant, (iv) the number of MC-generated MAH realizations used to compute observable-mass relationships.

5.2.5 Quantifying scaling relations

In our analysis of cluster scaling relations, we study individual, Monte Carlo-generated halo MAHs using the merger tree method of Parkinson et al. (2008). For each redshift of observation and cosmology considered, we generate 10,000 MAHs for haloes sampled uniformly in the mass range $12 \leq \log(M_{\rm vir}(z)/[h^{-1}M_\odot]) \leq 15.5$. For consistency with other studies, our analysis uses the mass range of $14 \leq \log(M_{\rm 200m}/[h^{-1}M_\odot]) \leq 15.6$ (a total of ~4,500 clusters), but we use the lower-mass systems to check for any dependence on the mass cutoff in the scaling relations (as well as study how properties of Γ depend on halo mass in Fig. 5.3). In the cluster mass regime, our assumption of a mass-independent $M_{\rm gas}(< r_{\rm ap})/M(< r_{\rm ap})$ ratio is well-

6. Note that the temperature is related to the effective temperature (Section 5.2.3) via $T(r) = [1 - f_{\rm nth}(r)]T_{\rm eff}(r)$.

justified (see the discussion in Sections 5.2.1 and 5.2.3). For each MAH, the $f_{\rm nth}(r)$ profile is evolved to the redshift of observation. Then, spanning a range of apertures, $r_{\rm ap}$, we compute the observables, $M_{\rm gas}(< r_{\rm ap})$, $T_{\rm mg}(< r_{\rm ap})$, and $Y_{\rm SZ}(< R_{\rm ap})$, and the corresponding halo mass, $M(< r_{\rm ap})$. We aim to elucidate how the slope, normalization, and scatter of the observable-mass relationships evolve with redshift and depend on aperture and cosmology. *Note that we use the same aperture to calculate both the observable and the enclosed mass.* It is sometimes the case in observational studies that the mass is measured within one aperture (e.g., $r_{\rm 500c}$) and the observable is measured within a larger aperture (e.g., $Y_{\rm SZ}[< 5R_{\rm 500c}]$), which can introduce additional effects due to the mass-concentration relation. We emphasize that in the limit that $f_{\rm nth} = 0$, the observables are computed purely from the KS01 model with $P_{\rm th} = P_{\rm tot}$, yielding the self-similar cluster scaling relations discussed in Section 5.2.1 with no scatter or deviation aside from that due to the mass-concentration relationship; thus, all scatter is due to the variance in the halo MAHs and its impact on the $f_{\rm nth}$ profile and halo concentrations. In particular, increased $f_{\rm nth}$ will result in $T_{\rm mg}$ and $Y_{\rm SZ}$ decreasing and falling below the self-similar curve.

For each observable, $X_{\rm obs}(< r_{\rm ap})$, we compute the best-fit power-law relationship

$$X_{\rm obs}(< r_{\rm ap}) = 10^{\alpha}\left(\frac{M(< r_{\rm ap})}{[10^{14}h^{-1}M_{\odot}]}\right)^{\beta}, \tag{5.19}$$

with α the normalization and β the power-law slope. We then compute the (natural) logarithmic residuals as

$$\mathcal{R} = \ln(X_{\rm obs,true}) - \ln(X_{\rm obs,fit}), \tag{5.20}$$

where $X_{\rm obs,fit}$ is computed from equation (5.19) given the $M(< r_{\rm ap})$ of each halo.[7]

We find that the ln-residuals for the $T_{\rm mg} - M$ and $Y_{\rm SZ} - M$ relations are *not* normally distributed due to a strong left-skew (i.e., there is a long tail towards large, negative $\mathcal{R}$). As we illustrate in Section 5.3.4, this is directly due to a right-skew in the recent MARs of the haloes and a correlation between Γ and non-thermal pressure support, which ultimately suppresses $Y_{\rm SZ}$. As shown in Fig. 5.3, the mean of and variance in Γ grows with both halo mass and redshift for MAHs generated via the Parkinson et al. (2008) method; this is directly responsible for a variety of trends in Section 5.3. Note that part of the strong right-skew is due to the fact that the MAR is bounded from below by zero, but is not bounded from above. A deviation from normality (and log-normality) of the residual distribution of $Y_{\rm SZ} - M$ in the form of a left-skew is also seen, albeit to a milder degree, in the non-radiative (NR) hydrodynamically-simulated clusters of Battaglia et al. (2012) (see their Fig. 22), indicating that in the absence of additional sources of non-thermal pressure beyond that

7. Note that our residual definition is opposite in sign to that which is normally used in the literature. As we show in Section 5.3.4, $\mathcal{R}$ as defined in equation (5.20) correlates with the halo MAR.

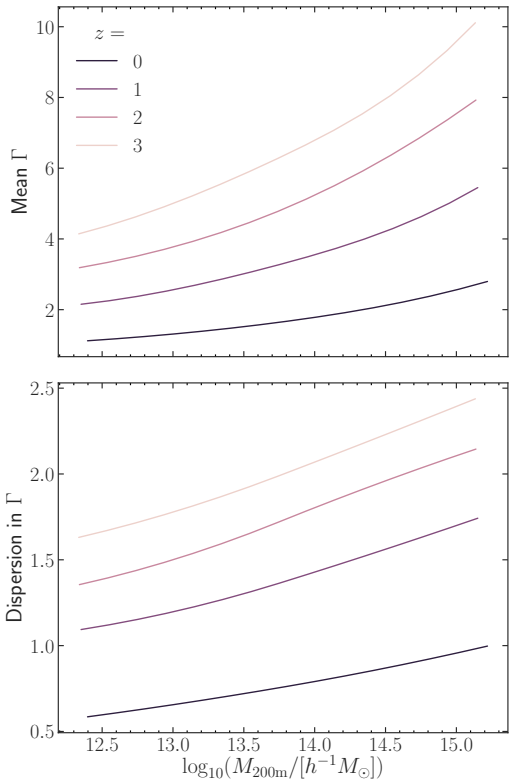

Figure 5.3: The mean of and dispersion in halo MARs, Γ (defined in equation [5.12]), as a function of halo mass and redshift for halo MAHs generated via the Parkinson et al. (2008) method. The Γ distribution is skewed (most strongly at low $M_{200\mathrm{m}}$ and z); hence, "dispersion" is defined as half of the $16-84$ percentile range. Note that both the mean of and dispersion in Γ grow with $M_{200\mathrm{m}}$ and z. These trends are ultimately responsible for the same trends seen in the non-thermal pressure fractions (Section 5.3.1) and for the increased scatter and decreased normalization in the scaling relations as z increases (Section 5.3.3).

introduced due to the halo assembly history, the residual distribution does indeed reflect the distribution of halo MARs. However, with the addition of radiative cooling, star formation, supernovae feedback, and AGN feedback, Battaglia et al. (2012) find that the residual distribution of $Y_{SZ} - M$ approaches normality (*not* log-normality). The relationship between the MAR and the $Y_{SZ} - M$ residuals will be discussed in more detail in Section 5.3.4.

The correlation between Γ and non-thermal pressure support also causes the scatter in the scaling relations to increase systematically with halo mass. Regardless of the ln-residual distribution's deviation from normality and heteroscedasticity, ordinary least squares remains the best linear unbiased estimator of the mass-observable regression coefficients (Plackett, 1950). These details, however, do affect how we should report the scatter seen in the observable-mass relations. Typically, in analysis of both simulated and observed clusters, the residual distribution is assumed to be log-normal with mean zero. Under this assumption, one can simply report the scatter as the standard deviation of the ln-residuals, $\sigma_{\mathcal{R}}$, which approximates the fractional/percent scatter, $\sigma_{\Delta X}/X$, to within 5% (10%) accuracy for $\sigma_{\mathcal{R}} \leq 0.1$ (0.2). Due to the substantial deviation from log-normality in our case and in an effort to make comparisons to results in the literature, we report scatter instead based on half of the $16 - 84$ percentile range of $\mathcal{R}$. Our reported percent scatters are smaller by roughly 1% (in absolute units, not relative) than they would be if we instead used the standard deviation of the ln-residuals.

5.3 Results

We start this section off by exploring the non-thermal pressure fraction profiles of the average cluster observed with a given mass at a particular redshift (Section 5.3.1). We then study the resulting average HSE mass bias introduced due to non-thermal pressure support (and its dependence on halo mass and redshift) in Section 5.3.2. We proceed to calculate the scaling relations of samples of individual clusters, studying their dependence on aperture radius, cosmology, redshift, and halo mass limit in Section 5.3.3. Lastly, we identify a strong correlation between the halo MAR and the $Y_{SZ} - M$ residual in Section 5.3.4, briefly discussing the potential utility of such a relationship.

5.3.1 Non-thermal pressure fractions

Since the cornerstone of our analysis is the SK14 model of the non-thermal pressure, we first study its predicted $f_{nth}(r)$ profiles for an average cluster of mass M_{200m} observed at z using the 'universal model' of the MAH from van den Bosch et al. (2014). In Fig. 5.4, we plot the $f_{nth}(r)$ profiles for clusters of several different masses as a function of r/r_{200m}. The choice of r_{200m} is motivated by Nelson et al. (2014b), who

z	0	1	2	3
$\nu_{200\mathrm{m}}$	$\log(M_{200\mathrm{m}}/[h^{-1}M_\odot])$			
1.16	13.00	11.61	10.30	9.11
2.01	14.20	13.13	12.14	11.27
4.10	15.40	14.60	13.88	13.27

Table 5.2: The peak heights studied in Fig. 5.4 and the corresponding halo masses at each redshift.

find that the $f_{\mathrm{nth}}(r/r_{200\mathrm{m}})$ profiles of their sample of NR hydrodynamically-simulated galaxy clusters is universal throughout their time evolution (this will be discussed more below); this universality is absent when normalized by $r_{200\mathrm{c}}$.

In the left panels of Fig. 5.4, we hold $M_{200\mathrm{m}}$ fixed and show how the f_{nth} radial profile changes with observation redshift. As halo mass increases, the non-thermal pressure fraction increases. This can be explained by the fact that higher mass haloes assemble at later times (e.g., Lacey & Cole, 1993; van den Bosch, 2002; Li et al., 2008); hence, their recent MAR will be higher than that of lower mass haloes (cf. Fig. 5.3). More non-thermal energy has been recently injected into a system with a higher recent MAR, which results in a larger f_{nth}. We also see that at fixed halo mass, f_{nth} is larger for clusters observed at higher redshift. This can be explained similarly to the previous point: in order for a halo to obtain a mass of M by $z_1 > z_2$, it must have accreted mass more rapidly than a halo with a mass of M at z_2 (cf. Fig. 5.3). Note that the fraction of non-thermal pressure is substantial, especially in the cluster outskirts — f_{nth} surpasses 50% by around $\sim r_{200\mathrm{m}}$ for high-mass haloes and haloes at large z.

We now explore the dependence of $f_{\mathrm{nth}}(r/r_{200\mathrm{m}})$ on peak height, $\nu_{200\mathrm{m}} = \delta_{\mathrm{c}}(z)/\sigma(M_{200\mathrm{m}})$.[8] In the right panels of Fig. 5.4, we hold $\nu_{200\mathrm{m}}$ fixed to several different values and show how $f_{\mathrm{nth}}(r/r_{200\mathrm{m}})$ evolves with redshift in each case (i.e., $M_{200\mathrm{m}}$ is varied with z such that $\nu_{200\mathrm{m}}$ remains constant). As is apparent, there is far less redshift evolution at fixed peak height than at fixed mass. We overplot the fitting function of Nelson et al. (2014b) for their universal $f_{\mathrm{nth}}(r/r_{200\mathrm{m}})$ profile, finding that for the peak height consistent with the $z = 0$ cluster masses studied in their work ($\nu_{200\mathrm{m}} \approx 4$), the predictions of the SK14 model agree well with what is seen in the simulations. There is an exception to this agreement, however, in the central regions of the clusters, where the model underpredicts the non-thermal pressure fraction compared to that seen in the Nelson et al. (2014b) simulations. As discussed in Shi et al. (2015), this is likely due to (i) the model's assumption of a one-to-one relationship between the cluster radius and the turbulence dissipation timescale (note that this assumption is the primary source of the f_{nth} radial dependence) and (ii) the potential need to incorporate radius and redshift dependence into η to properly model the relative importance of high-Mach accretion shocks and low-Mach internal shocks. Recently, Shi et al. (2018) found that the turbulence

8. Here, $\delta_{\mathrm{c}}(z) = \delta_{\mathrm{c}}(z=0)/D_+(z)$ is the critical overdensity for collapse (Gunn & Gott, 1972), $D_+(z)$ is the linear growth factor normalized to unity at $z = 0$, and $\sigma(M_{200\mathrm{m}})$ is the RMS mass fluctuation in a Lagrangian volume corresponding to $M_{200\mathrm{m}}$.

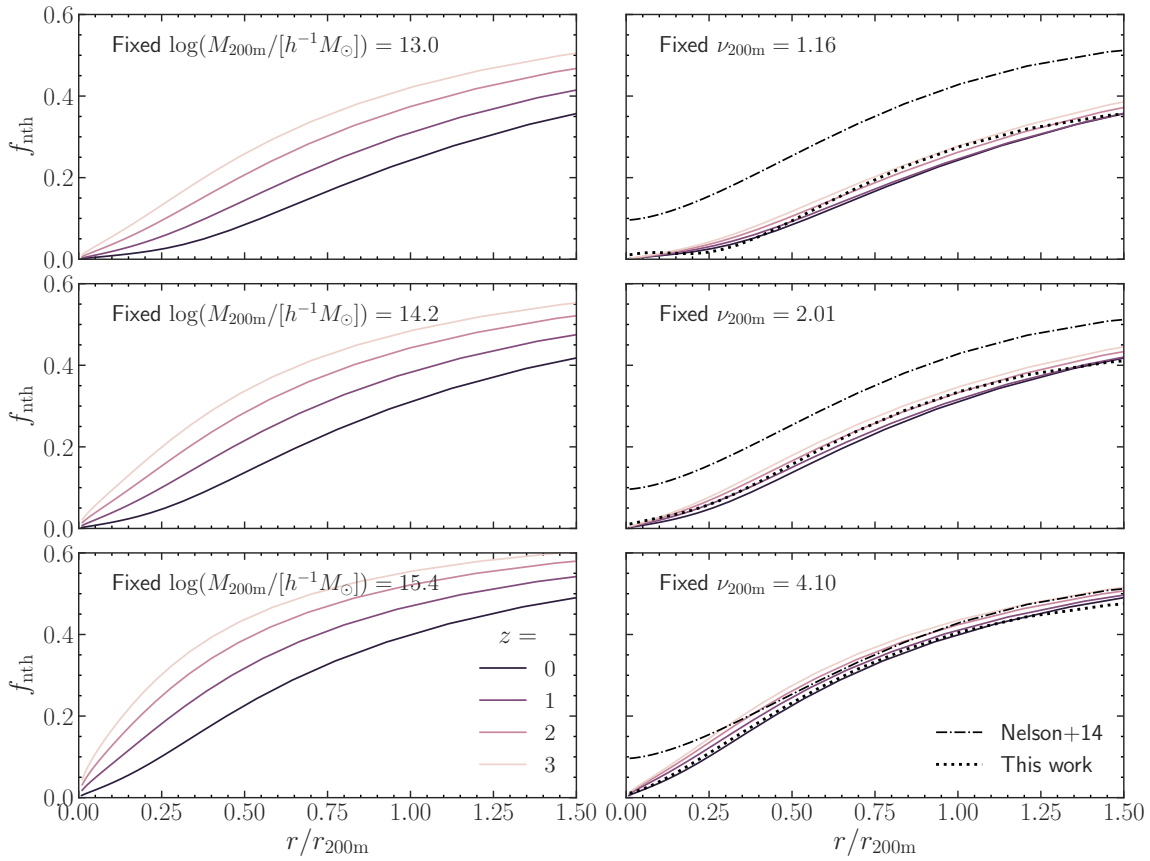

Figure 5.4: The non-thermal pressure fraction profiles of clusters, $f_{\mathrm{nth}}(r/r_{200\mathrm{m}})$, as predicted by the SK14 model combined with the 'universal model' of the MAH from van den Bosch et al. (2014). (*left*) Each panel holds the cluster mass, $M_{200\mathrm{m}}$, fixed and varies the redshift of observation. As either $M_{200\mathrm{m}}$ or z increases, the non-thermal pressure fraction increases due to the increased recent mass accretion rate. (*right*) Each panel holds the peak height, $\nu_{200\mathrm{m}}$, fixed such that the $z = 0$ mass is the same as that in the corresponding left panel. The masses corresponding to each peak height at the different redshifts are listed in Table 5.2. There is minimal redshift evolution in $f_{\mathrm{nth}}(r/r_{200\mathrm{m}})$ at fixed peak height. The 'universal profile' seen in the simulated clusters of Nelson et al. (2014b) is over-plot (dot-dashed line), illustrating the peak height-dependence that was not seen in their cluster sample due to their limited $z = 0$ mass range. Our fitting function described by equation (5.21) and Table 5.3 (dotted line) incorporates $\nu_{200\mathrm{m}}$-dependence and reproduces the SK14 model at $z = 1$ at roughly 10% accuracy in the radial range of $0.2 \leq r/r_{200\mathrm{m}} \leq 2.0$.

dissipation timescale measured in simulations is indeed shorter at smaller cluster radii and suggested that this is the case due to the stronger density stratification in the cluster core. Following this, Shi & Zhang (2019) confirmed the role of the density stratification and indicated that the buoyancy time ($t_{\rm BV}(r)$; i.e., the inverse of the Brunt-Väisälä frequency) may more accurately capture the timescale of turbulence dissipation. The buoyancy time is nearly the same as the orbital time outside of the cluster core; however, for non-cool-core clusters, the core in the cluster entropy profile results in a core in the buoyancy time profile (i.e., $t_{\rm BV}$ approaches a constant, non-zero value as $r \to 0$). We find that using $t_{\rm BV}$ for the turbulence dissipation timescale causes the $f_{\rm nth}$ profiles to change by less than 10% outside of the cluster core region ($\lesssim 0.2 r_{\rm 200m}$) relative to the fiducial case of $t_{\rm orb}$ — this propagates to a difference of only a few percent in our subsequent $Y_{\rm SZ}$ model predictions, since the signal is dominated by the cluster outskirts.

The model does not predict a universal $f_{\rm nth}(r/r_{\rm 200m})$ profile, which clearly has a dependence on $\nu_{\rm 200m}$ that, to good approximation, accounts for the dependence on both $M_{\rm 200m}$ and z. At first, this appears to be at odds with the simulated clusters studied in Nelson et al. (2014b). However, they studied the evolution of a cluster sample through *time*, with $M_{\rm 200m}$ only spanning half an order of magnitude in the range $14.8 < \log(M_{\rm 200m}/[h^{-1}M_\odot]) < 15.4$ at $z = 0$. This detail, combined with their use of a z-dependent mass cutoff for the cluster sample, likely resulted in the Nelson et al. (2014b) sample spanning an insufficient range in $\nu_{\rm 200m}$ to isolate evolution in redshift from universality in $f_{\rm nth}(r/r_{\rm 200m})$ at fixed $\nu_{\rm 200m}$.

Motivated by our finding that, to good approximation, $f_{\rm nth} = f_{\rm nth}(r/r_{\rm 200m}|\nu_{\rm 200m})$, we present a fitting function for the non-thermal pressure fraction that includes this $\nu_{\rm 200m}$-dependence:

$$f_{\rm nth}(\tilde{r}|\nu) = 1 - \left[A\left(1 + e^{-(\tilde{r}/B)^C}\right)\left(\frac{\nu}{4.1}\right)^{\frac{D}{(1+[\tilde{r}/E]^F)}}\right]. \tag{5.21}$$

Here, $\nu = \nu_{\rm 200m}$ and $\tilde{r} = r/r_{\rm 200m}$. The parameters of this function are calibrated to match the $z = 1$ predictions of the model across $1.0 \leq \nu_{\rm 200m} \leq 4.2$ and are listed in Table 5.3. The fit, shown as dotted curves in Fig. 5.4, is accurate to roughly 10% over the radial range of $0.2 \leq r/r_{\rm 200m} \leq 2.0$. In a future work, we will further explore this $\nu_{\rm 200m}$-dependence and the sensitivity of the $f_{\rm nth}(r/r_{\rm 200m})$ predictions to cosmology and more realistic definitions of the turbulence dissipation timescale.

5.3.2 Hydrostatic mass bias

As discussed in the introduction, cluster mass inferences based on X-ray and SZ observations are typically made under the assumption of HSE between the observed thermal pressure profile and the gravitational potential. The true cluster mass, however, is related to the *total* pressure profile, and thus any unaccounted-for sources of non-thermal pressure result in underprediction of the cluster mass. The $f_{\rm nth}$ profiles predicted

Parameter	Value
A	0.495
B	0.719
C	1.417
D	−0.166
E	0.265
F	−2.116

Table 5.3: Calibrated parameters of the $f_{\rm nth}(r/r_{\rm 200m}, \nu_{\rm 200m})$ fitting function described by equation (5.21), which reproduces the model non-thermal pressure fractions to roughly 10% accuracy in the radial range of $0.2 \leq r/r_{\rm 200m} \leq 2.0$ at $z = 1$. Note that there is only a weak redshift dependence in the model predictions, as can be seen in Fig. 5.4, so this fitting function can be easily used to make rough predictions regardless of redshift.

by the SK14 model can be used to estimate the corresponding HSE mass bias.

From the HSE equation (i.e., equation [5.3]), one can compute how much the true mass, M, is underpredicted ($M^{\rm HSE}$) as a function of mass and redshift. Assuming an accurate determination of the gas density and thermal pressure profiles, which can be made possible through the combination of X-ray and SZ observations (e.g., Ameglio et al., 2009; Eckert et al., 2019; Ettori et al., 2019), this underprediction is written as

$$\frac{M^{\rm HSE}(<r)}{M(<r)} = \frac{{\rm d}P_{\rm th}/{\rm d}r}{{\rm d}P_{\rm tot}/{\rm d}r} = [1 - f_{\rm nth}(r)] - P_{\rm tot}(r)\frac{{\rm d}f_{\rm nth}/{\rm d}r}{{\rm d}P_{\rm tot}/{\rm d}r}. \tag{5.22}$$

Since ${\rm d}P_{\rm tot}/{\rm d}r$ is negative and ${\rm d}f_{\rm nth}/{\rm d}r$ is positive, this ratio should always be larger than $1 - f_{\rm nth}(r)$ for measurements of mass enclosed within r. Note that this estimate of the HSE bias neglects potential effects due to the deviation from spherical symmetry and projection effects. In Fig. 5.5, we plot the predictions for $M^{\rm HSE}_{\rm 500c}/M_{\rm 500c}$ as a function of $M_{\rm 500c}$ and redshift of observation. We use $r_{\rm ap} = r_{\rm 500c}$ ($\approx 0.4 r_{\rm 200m}$) since this is the aperture most commonly used for X-ray-based cluster mass estimation. At this radius, the SK14 model is in good agreement with the $f_{\rm nth}$ profiles of the simulated clusters of Nelson et al. (2014b), which only include NR hydrodynamics. Hence, additional sources of non-thermal pressure due to magnetic fields, cosmic rays, supernova feedback, among others, are not included and thus, we expect these estimates of the magnitude of the HSE bias to be *lower bounds*. The SK14 model predicts that the magnitude of the HSE bias increases considerably with cluster mass and observation redshift. At $z = 0$, HSE-based masses underestimate the true masses by less than 10% even for the highest mass clusters. However, at $z \sim 2 - 3$, the HSE bias results in substantial underprediction of the true mass, by roughly 20% at group scales and as much as 30 – 40% for high-mass clusters.

In the $z = 0.25$ simulated cluster sample from the *BAHAMAS* (McCarthy et al., 2017) and *MACSIS* (Barnes et al., 2017a) hydrodynamic simulations studied in Henson et al. (2017), which include star formation, radiative cooling, and feedback from supernovae and AGN (hereafter referred to as "full-physics" simulations),

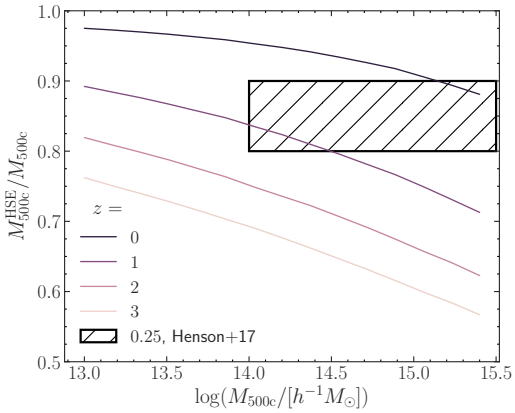

Figure 5.5: The HSE mass bias for $M_{\rm 500c}$, computed using equation (5.22), as predicted using the SK14 model of the non-thermal pressure. The "bias" increases as $M^{\rm HSE}_{\rm 500c}/M_{\rm 500c}$ decreases. These results are roughly consistent with the simulated clusters studied in Henson et al. (2017, at $z = 0.25$, hatched black box), although our results can be considered lower bounds, as sources of non-thermal pressure in addition to those due to mass assembly are not modeled and will increase the bias. The mass bias increases substantially with redshift, motivating further simulation studies focused on the HSE bias redshift evolution.

the HSE bias found is $M^{\rm HSE}_{\rm 500c}/M_{\rm 500c} \approx 0.8 - 0.9$ in the mass range $14 \leq \log(M_{\rm 500c}/[h^{-1}M_\odot]) \leq 15.5$. Their bias is only marginally larger than that predicted by our model (Fig. 5.5), which is most likely due to the additional sources of non-thermal pressure captured in the full-physics simulations. Similarly, Ansarifard et al. (2020) reports a median of $M^{\rm HSE}_{\rm 500c}/M_{\rm 500c} \approx 0.9$ for simulated clusters at $z = 0$ in a similar mass range. On the other hand, using synthetic X-ray observations, Barnes et al. (2021) report biases as significant as $M^{\rm HSE}_{\rm 500c}/M_{\rm 500c} \approx 0.7$ for simulated clusters at $z = 0.1$ with $\log(M_{\rm 500c}/M_\odot) \approx 15.3$ — they find that this is primarily due to the use of a single temperature fit to the full cluster spectrum. Additionally, Hurier & Angulo (2018) report a similar bias of 0.73 ± 0.07 when using CMB lensing to estimate cluster masses, although they find no significant dependence on mass or redshift. For reference, a larger bias is necessary ($M^{\rm HSE}_{\rm 500c}/M_{\rm 500c} \approx 0.6$) in order to resolve the tension between cosmological parameter estimates based on the cluster mass function and cosmic microwave background approaches (Salvati et al., 2019).

We emphasize that these calculations are based on the average MAH for a cluster observed with a given mass and redshift. Clusters that are more disturbed (i.e., have a higher recent MAR) will generally have larger biases than the average, as their non-thermal pressure fraction will be larger. In fact, the difference between the HSE bias of an individual cluster and the average (at fixed halo mass and redshift) should correlate with the MAR; as we discuss in Section 5.3.4, a strong correlation also arises between the residuals of the $Y_{\rm SZ} - M$ relation and the MAR. To date, we are not aware of any simulation studies that characterize the evolution of the HSE bias over a reasonably large range of redshifts. Based on the results of Fig. 5.5, such a study is warranted, as the redshift dependence of the HSE bias predicted will be important to account for

in cluster count analyses that include high-z cluster samples from future surveys.

5.3.3 Cluster scaling relations

Having demonstrated that the average $f_{\rm nth}$ profiles (and resultant HSE biases) predicted by the SK14 model are in good agreement with predictions from hydrodynamical simulations, we proceed to use the model to study the cluster scaling relationships. In Fig. 5.6, we plot the best fit normalization, slope, and percent scatter for the $z = 0$ relations as a function of $r_{\rm ap}$. In order to provide insight into the model predictions and disentangle the nonlinear interactions between its various components, we calculate the cluster observables in three different ways. First, we compute cluster observables using the "full model" described in Sections 5.2.3 and 5.2.4. We then repeat the calculations while holding the halo concentrations fixed to $c_{\rm vir} = 5$ (referred to as the "fixed $c_{\rm vir}$ model"), isolating the impact of the mass-concentration relation. Going further, we perform a third set of calculations: while continuing to hold $c_{\rm vir}$ fixed, we now also replace the radius-dependent turbulence dissipation timescale with its value at $r_{\rm 200m}$ (i.e., $t_{\rm dis}(r) = t_{\rm dis}(r_{\rm 200m})$; referred to as the "fixed $c_{\rm vir}$ and $t_{\rm dis}$ model"). This elucidates the impact of the radial dependence of $t_{\rm dis}(r)$. Note that in this final case, the $f_{\rm nth}$ profile is nearly constant with radius for a given halo and all variation in $f_{\rm nth}$ between haloes is due to variation in MAHs.

5.3.3.1 $M_{\rm gas} - M$ relation

Beginning with the $M_{\rm gas} - M$ relation (middle column of Fig. 5.6), our model predicts no scatter in the absence of a MAH-dependence on the concentration. This is simply due to our use of the KS01 model for $\rho_{\rm gas}$, which has no dependence on the halo MAH or $f_{\rm nth}$ but only on $c_{\rm vir}$ (the implications of this are discussed in more detail in Sections 5.3.3.3 and 5.3.3.6). In the full model, the scatter goes to zero and the slope goes to unity as $r_{\rm ap} \rightarrow 2r_{\rm 200m}$; this is simply due to our chosen gas density normalization that $M_{\rm gas}(< 2r_{\rm 200m}) = (\Omega_{\rm b}/\Omega_{\rm m})M(< 2r_{\rm 200m})$. The effect of the mass-concentration relation and its intrinsic scatter on the shape of both the dark matter and gas density profiles is responsible for the small change in slope (and increase in scatter) of $M_{\rm gas} - M$ as aperture radius decreases (see e.g., Fujita & Aung, 2019). Even with fixed concentrations, the difference between the gas and dark matter density profile shapes is responsible for a slight dependence on $r_{\rm ap}$ in the $M_{\rm gas} - M$ normalization.

5.3.3.2 $T_{\rm mg} - M$ relation

Next, we direct our attention to the mass-weighted temperature, $T_{\rm mg}$ (left-hand column of Fig. 5.6). In the fixed $c_{\rm vir}$ and $t_{\rm dis}$ model, the scatter and slope are independent of aperture. More-massive clusters have

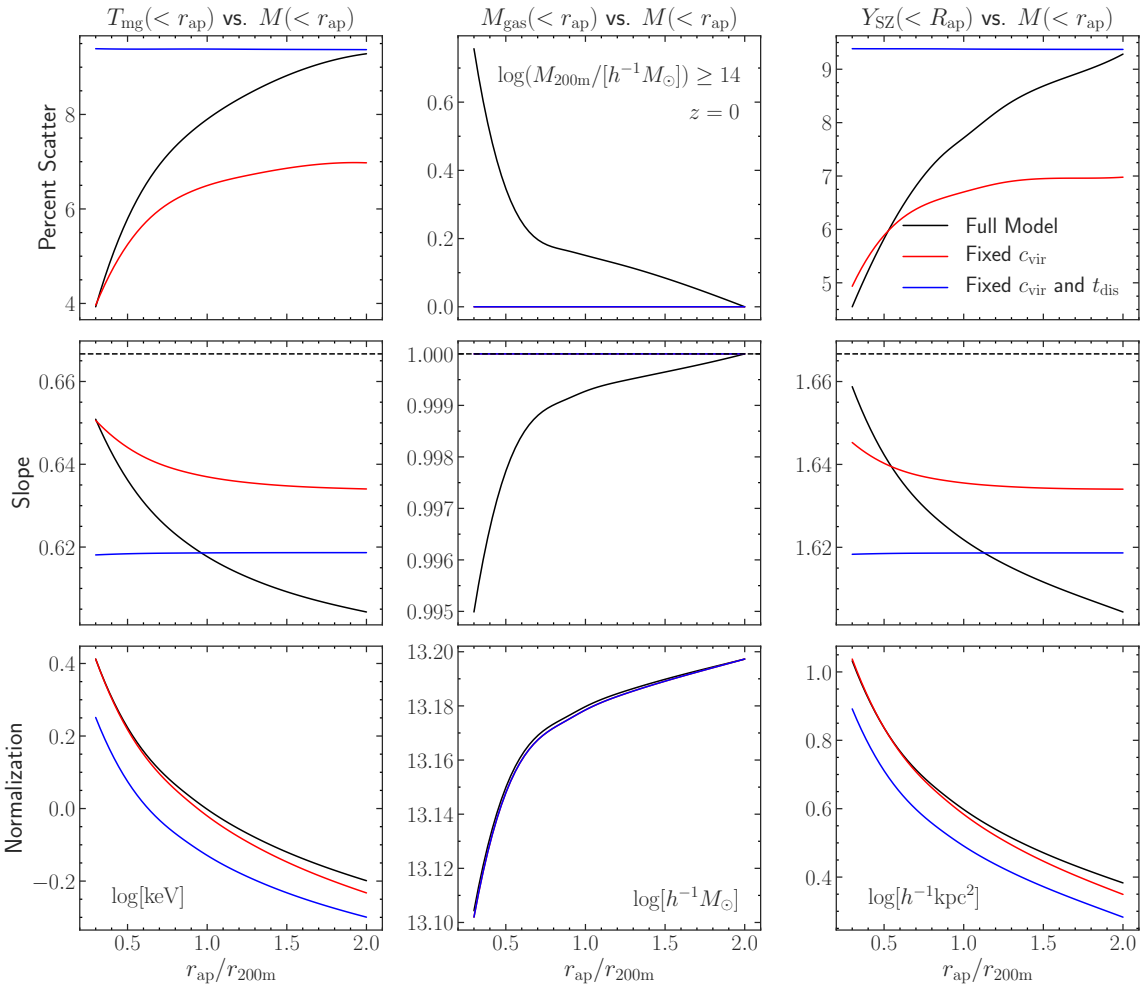

Figure 5.6: The best fit normalization, slope, and percent scatter (i.e., half of the 16−84 percentile interval of $\mathcal{R}$) of the $z = 0$ observable-mass relations, described by equation (5.19) for $T_{\rm mg}(< r_{\rm rap}) - M(< r_{\rm ap})$, $M_{\rm gas}(< r_{\rm rap}) - M(< r_{\rm ap})$, and $Y_{\rm SZ}(< R_{\rm ap}) - M(< r_{\rm ap})$. The fit parameters are shown as a function of the aperture radius, $r_{\rm ap}$, in units of $r_{\rm 200m}$; note that the same aperture is used to compute both the observable and the total mass. In these fits, ∼4,500 clusters in the mass range $14 \leq \log(M_{\rm 200m}/[h^{-1}M_\odot]) \leq 15.6$ are used. The black dashed lines indicate the slopes predicted by the self-similar relations. The observables are computed using the "full model" described in Sections 5.2.3 and 5.2.4 (black curves) as well as two simplified models, one of which holds $c_{\rm vir} = 5$ fixed (red curves) in order to isolate the effects of the mass-concentration relation and another that holds both $c_{\rm vir} = 5$ fixed *and* replaces the radius-dependent turbulence dissipation timescale with its value at $r_{\rm 200m}$ (blue curves), isolating the interaction between $r_{\rm ap}$ and the radius-dependence of $f_{\rm nth}(r)$. See the main text in Section 5.3.3 for detailed explanations of the trends with $r_{\rm ap}$.

larger MAR, which drives larger $f_{\rm nth}$ (see the left panels of Fig. 5.4). For $f_{\rm nth}$ independent of radius, we have that $T_{\rm mg} \propto (1 - f_{\rm nth})$, which, combined with the fact that $f_{\rm nth}$ grows with halo mass, results in the slope of $T_{\rm mg} - M$ lying below that of self-similarity. In this simplified model, the aperture-independent scatter in $T_{\rm mg} - M$ is also driven solely by variation in halo MAHs and is most sensitive to the mass evolution over the previous dynamical time. Moving on to the fixed $c_{\rm vir}$ model, we notice that incorporation of a radius-dependent $t_{\rm dis}(r)$ introduces a dependence on aperture into the slope and scatter of $T_{\rm mg} - M$. Higher mass clusters tend to have $f_{\rm nth}(r)$ profiles that are overall larger in magnitude *and* grow more rapidly with radius (most notably in the inner radii; see once again the left panels of Fig. 5.4). Hence, their temperature profiles will be more suppressed overall relative to self-similarity. In addition, since $\mathrm{d}f_{\rm nth}(r)/\mathrm{d}r$ increases with halo mass (in the inner radii), the slope of $T_{\rm mg} - M$ decreases further from self-similarity as $r_{\rm ap}$ increases. The scatter in $f_{\rm nth}(r)$ grows with radius due to the radially increasing $t_{\rm dis}(r)$; because of this, the scatter in $T_{\rm mg} - M$ grows with aperture radius. Lastly, by looking at the full model, we see two effects due to the mass-concentration relation. First, $c_{\rm vir}(M, z)$ results in further reduction in the $T_{\rm mg} - M$ slope away from self-similarity. Additionally, the variance in $c_{\rm vir}(M, z)$ propagates to additional scatter in $T_{\rm mg} - M$ that becomes more substantial as $r_{\rm ap}$ increases. Finally, the normalization of the $T_{\rm mg} - M$ relation decreases with increasing aperture for a simple reason. Since the cluster temperature decreases with radius, the mass-weighted temperature must decrease as the aperture radius increases. In addition, the pivot mass used for the relations is $M(< r_{\rm ap}) = 10^{14} h^{-1} M_\odot$ regardless of aperture. Hence, this pivot mass at larger $r_{\rm ap}$ corresponds to a smaller total (virial) mass and thus a lower temperature normalization.

5.3.3.3 $Y_{\rm SZ} - M$ relation

Lastly, we turn to the integrated SZ signal, $Y_{\rm SZ}$ (right-hand column of Fig. 5.6). Since $Y_{\rm SZ}$ is simply the cylindrically-integrated pressure profile, to good approximation $Y_{\rm SZ} \propto M_{\rm gas} T_{\rm mg}$. This relationship bares out straightforwardly in Fig. 5.6, as the slope of the $Y_{\rm SZ} - M$ relation evolves roughly as the sum of the slopes of the $T_{\rm mg} - M$ and $M_{\rm gas} - M$ relations. For the smallest values of $r_{\rm ap}$, we have verified that the slight disagreement between the slope of $Y_{\rm SZ} - M$ and the sum of the $T_{\rm mg} - M$ and $M_{\rm gas} - M$ slopes is simply due to projection effects that manifest due to the different impact of halo concentrations on spherically- and cylindrically-integrated quantities. The clusters from the NR hydrodynamics simulations of Stanek et al. (2010) yield a $Y_{\rm SZ}(< r_{\rm 200c}) - M_{\rm 200c}$ relation slope of 1.651 ± 0.003. For comparison, and noting that $r_{\rm 200c} \approx 0.6 r_{\rm 200m}$, we find a slope in $Y_{\rm SZ}(< R_{\rm 200c}) - M_{\rm 200c}$ of roughly 1.635 (note that this reduces slightly to 1.63 if we instead compute $Y_{\rm SZ}(< r_{\rm 200c}) - M_{\rm 200c}$ with a spherically-integrated $Y_{\rm SZ}$, which is not shown). In our calculations, the only source of scatter in $M_{\rm gas} - M$ is the mass-concentration relation. However, as

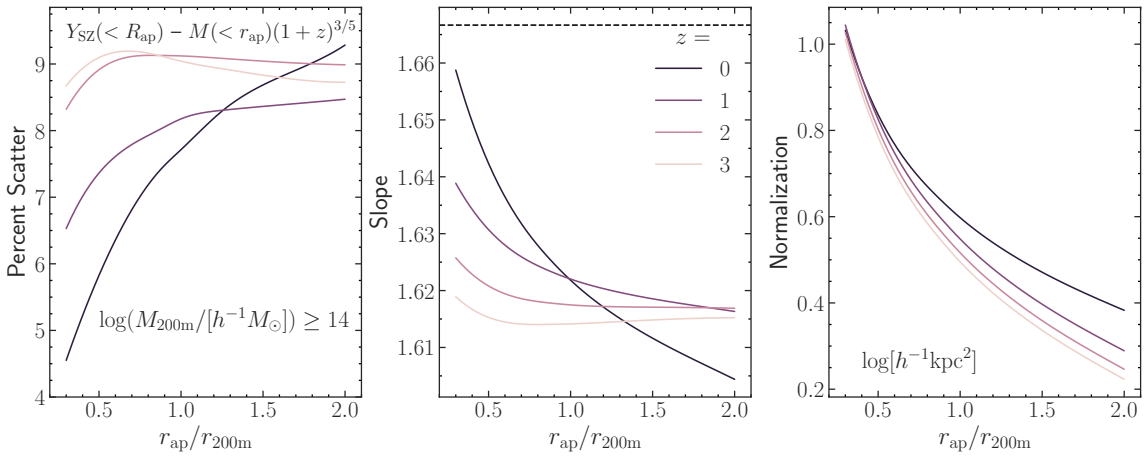

Figure 5.7: The best fit normalization, slope, and percent scatter of the $Y_{SZ}(< R_{ap}) - M(< r_{ap})$ relation, with masses scaled by the self-similarity evolution factor, $(1+z)^{3/5}$. The fit parameters are shown as a function of r_{ap} and different curves illustrate the redshift evolution from $z = 0$ to $z = 3$. In these fits, ~4,500 clusters uniformly distributed in the mass range $14 \leq \log(M_{200m}(z)/[h^{-1}M_\odot]) \leq 15.6$ are used for each z. As observation redshift increases, the slope and normalization tend to decrease while the scatter increases. The interaction of the redshift-dependence of the mass-concentration relation is responsible for the apparent trend-reversals around $r_{ap} \approx r_{200m}$. Similar results have been seen in hydrodynamics simulations (e.g., Nagai, 2006; Battaglia et al., 2012; Le Brun et al., 2017; Planelles et al., 2017).

described above, the variance in the cluster MAHs drives the scatter in $T_{mg} - M$ and increases considerably with aperture. Thus, in our model, the scatter in the $Y_{SZ} - M$ relation is driven predominantly by the scatter in the $T_{mg} - M$ relation. A more realistic model of the gas density profile that incorporates additional baryonic processes will introduce additional variance into $M_{gas} - M$, as well as stronger covariance between M_{gas} and T_{mg} (see e.g., Stanek et al., 2010), which will ultimately increase the scatter in $Y_{SZ} - M$. Hence, our scatter estimates should be regarded as lower bounds (see additional discussion in Section 5.3.3.6). Regarding the reduction in normalization and slope of $Y_{SZ} - M$ with increasing aperture, we find similar trends to those reported in Nagai (2006).

5.3.3.4 Redshift evolution

Having explored the non-linear interactions between aperture radius, halo concentration, and turbulence dissipation timescales in our model, we move on to study the redshift evolution of $Y_{SZ} - M$. In Fig. 5.7, we plot the best fit normalization, slope, and percent scatter for the $Y_{SZ} - M$ scaling relation as a function of r_{ap} for different samples of clusters observed at $0 \leq z \leq 3$. When using a spherical overdensity definition relative to the mean matter density (such as r_{200m}), the full self-similar scaling relation is $Y_{SZ} \propto M^{5/3}(1+z) = [M(1+z)^{3/5}]^{5/3}$ (see Section 5.2.1). Thus, scaling the masses by $(1+z)^{3/5}$ accounts for the redshift evolution predicted by the self-similar model. Any additional redshift evolution in the normalization or

slope of $Y_{SZ}-M(1+z)^{3/5}$ indicates z-dependent deviations from self-similarity. The model predicts some rich trends with observation redshift. The normalization of $Y_{SZ}-M(1+z)^{3/5}$ decreases slightly with increasing z, with the decrease being larger when a larger r_{ap} is used. This is simply due to the fact that at earlier times, halo MARs were generally higher (see Fig. 5.3), resulting in an overall increase in non-thermal pressure support due to turbulence, and thus suppression in Y_{SZ}, with increasing z (as in Fig. 5.4). For $r_{ap} \lesssim r_{200m}$, the slope of the relation decreases with increasing z. This is due to the fact that f_{nth} in the inner regions increases more strongly with z in more massive haloes (this can be seen in the left panels of Fig. 5.4). The apparent trend-reversal at larger aperture radii is caused by the mass-concentration relation and its redshift evolution. The model also predicts that scatter in $Y_{SZ} - M$ increases with z, which is directly a consequence of the increased variance in halo MARs at earlier times (see Fig. 5.3).

These redshift evolution trends are in overall agreement with predictions from NR hydrodynamical simulations, most clearly with regards to the scatter evolution. The studies by Battaglia et al. (2012), Le Brun et al. (2017), and Planelles et al. (2017) all find that the scatter in $Y_{SZ} - M$ increases slightly with increasing z in their NR simulations (although only for high-mass clusters in the case of Le Brun et al. 2017). On the other hand, only Battaglia et al. (2012) finds that the $Y_{SZ} - M$ slope tends to decrease slightly away from self-similarity with z, whereas Le Brun et al. (2017) and Planelles et al. (2017) find minimal redshift evolution in the slope. The $Y_{SZ} - M$ slope increases when going to the full-physics AGN simulations slightly in Battaglia et al. (2012) and significantly (up to $\sim$2) in Le Brun et al. (2017), whereas it remains virtually unchanged in Planelles et al. (2017), highlighting a point of tension between simulation results. These studies (as well as Nagai, 2006) have reported that the redshift evolution of the normalization shows no significant deviation from self-similarity when $r_{ap} = r_{500c}$, consistent with our findings for $r_{ap} \approx 0.4 r_{200m}$. However, the predictions of Fig. 5.7 show that deviations from self-similarity are expected to increase in magnitude when larger aperture radii are employed. This, combined with the current tension between the results of various simulation studies (particularly with regards to the dependence of the $Y_{SZ} - M$ slope on z and AGN physics), suggests that the redshift evolution (and its dependence on r_{ap}) of cluster scaling relations needs to be studied in more depth using large cluster counts. In particular, a comparison between NR and full-physics simulations will help determine whether or not the trends due to variance in MAHs predicted by our model are washed out by additional physical processes (such as AGN and supernova feedback, etc.). With upcoming surveys pushing to larger cluster counts and higher z, characterizing the redshift evolution of these scaling relations is of paramount importance. If our model prediction that scatter in the relations increases significantly with redshift is correct, then it will be important to continue to develop lower-scatter mass proxies with less sensitivity to redshift in order to maximally utilize upcoming high-redshift cluster data to their full potential for precision cosmology.

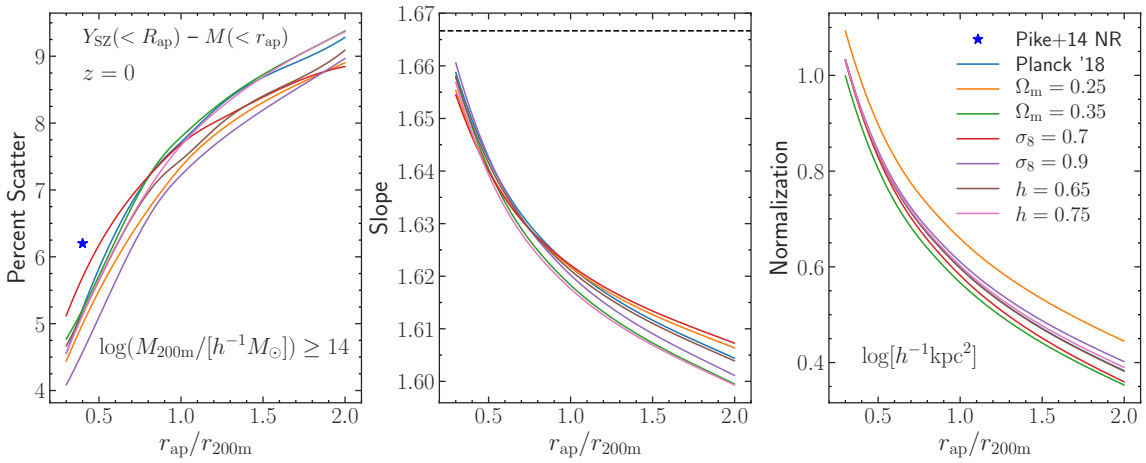

Figure 5.8: The best fit normalization, slope, and percent scatter of the $Y_{SZ}(< R_{ap}) - M(< r_{ap})$ scaling relation at $z = 0$. The fit parameters are shown as a function of the aperture radius, r_{ap}, in units of r_{200m}. In these fits, ~4,500 clusters in the mass range $14 \leq \log(M_{200m}/[h^{-1}M_\odot]) \leq 15.6$ are used. Each curve represents a different cosmology, varied about the fiducial Planck Collaboration et al. (2020) cosmological parameters — the variations in $Y_{SZ} - M$ with cosmology are subtle, but the trends we find are consistent with the simulations of Singh et al. (2020). Our predicted scatter in $Y_{SZ} - M$ for $r_{ap} = r_{500c} \approx 0.4 r_{200m}$ (at $z = 0$) is only slightly below the scatter seen in the NR hydrodynamical simulations of Pike et al. (2014). The majority of simulation studies predict scatter in the range of 10 – 16%. Hence, much of the scatter in $Y_{SZ} - M$ is simply due to inter-cluster variation in the mass assembly histories, which drives variance in the cluster $f_{nth}(r)$ profiles.

5.3.3.5 Dependence on cosmology and halo mass cutoff

In Fig. 5.8, we consider the impact of single-parameter variations about the fiducial (Planck Collaboration et al., 2020) cosmology on $Y_{SZ} - M$ at $z = 0$. Over the range of cosmologies studied, we find small but systematic trends. Recently, Singh et al. (2020) studied the effect of variations in the cosmological parameters on X-ray-based cluster scaling relations using full-physics hydrodynamics simulations based on the *Magneticum* suite.[9] Using an aperture of $r_{ap} = r_{vir}$, they find that the slope and normalization of $Y_{SZ} - M$ systematically decrease with increasing Ω_m. We qualitatively reproduce these trends. While the changes to the properties of the $Y_{SZ} - M$ relation due to large changes in the cosmological parameters (relative to the posterior distributions of the Planck Collaboration et al. 2020 parameters) are small, both the present work and Singh et al. (2020) illustrate that more accurate models of the cluster scaling relations (and their dependence on cosmology) may eventually provide an additional approach to constraining the cosmological parameters given large ($\sim 10^4 - 10^5$) cluster samples from next-generation missions, such as *eROSITA*, Simons Observatory, and CMB-S4. However, for such an approach to be feasible, future analytical gas models must account for additional significant physical processes (see Section 5.4) and the accuracy of their predictions

9. http://magneticum.org/

must be validated against simulations that span a realistic range of cosmological parameters.

The level of variation in the slope and scatter of the $Y_{SZ} - M$ relation caused by changes in the cosmology are similar in magnitude to the level of variation imparted due to changing the minimum halo mass cutoff. In particular, decreasing the minimum halo mass used to compute these relations tends to decrease the overall scatter in the $Y_{SZ} - M$ relation, since the dispersion in Γ is lower for low-mass haloes (see Fig. 5.3). In addition, a lower mass limit tends to move the slope closer towards self-similarity due to the decrease in f_{nth} with decreasing halo mass. We find that decreasing the mass cutoff from $\log(M_{200m}/[h^{-1}M_\odot]) \geq 14$ to $\log(M_{200m}/[h^{-1}M_\odot]) \geq 12$ decreases the percent scatter in $Y_{SZ}(< R_{ap}) - M(< r_{ap})$ by ~1% (in absolute units, not relative) and increases its slope by ~0.01 towards self-similarity. While these changes are small, this does impart a degeneracy between the mass regime used and the cosmological parameters that may become important with sufficiently large cluster samples. Importantly, we emphasize that our model does not include various physical sources of non-thermal pressure support that become increasingly important for low-mass haloes (e.g., feedback). In addition, it remains unclear how valid our choice of gas density normalization (i.e., fixed $M_{gas}(< 2r_{200m})/M(< 2r_{200m}) = \Omega_b/\Omega_m$) is for low-mass haloes. Hence, these trends with respect to the halo mass cutoff must be considered with reservation. The halo mass-dependence of f_{nth} results in a more complex relation between halo mass and observable than a simple power law can capture; future analyses should consider employing localized linear regression (e.g., as used in Farahi et al., 2018; Anbajagane et al., 2020) in order to quantify the mass-dependence of the scaling relation properties.

5.3.3.6 Scatter comparison with simulations and observations

The simple model of SK14 demonstrates that a substantial fraction of the total scatter in the $Y_{SZ} - M$ relation is likely to arise from inter-cluster variance in the non-thermal pressure, which in turn arises from variance in the halo MAHs. There have been numerous studies that address the scatter in the $Y_{SZ} - M$ relation using simulations (da Silva et al., 2004; Nagai, 2006; Stanek et al., 2010; Battaglia et al., 2012; Kay et al., 2012; Sembolini et al., 2013; Pike et al., 2014; Yu et al., 2015; Hahn et al., 2017; Le Brun et al., 2017; Planelles et al., 2017; Henden et al., 2019; Singh et al., 2020) as well as observations using both weak-lensing and HSE X-ray masses (Bonamente et al., 2008; Hoekstra et al., 2012; Marrone et al., 2012; Planck Collaboration et al., 2014b; Czakon et al., 2015; Sereno et al., 2015; Nagarajan et al., 2019). At $z = 0$, most simulation studies find an intrinsic scatter in Y_{SZ} at fixed mass of 10−16% when using $R_{ap} = R_{500c} \approx 0.4R_{200m}$. There are indications in these studies that the scatter increases slightly when going from NR runs to full-physics simulations with AGN (see e.g., Battaglia et al., 2012). On the lower end, the $Y_{SZ}(< R_{500c}) - M(< r_{500c})$ relation computed using the NR simulations of Pike et al. (2014) has a scatter of just 6%. For comparison, the intrinsic scatter in the

$Y_{SZ}-M$ relation predicted by our model, using the same aperture (see Fig. 5.8) and with the same mass cutoff ($\log(M_{200m}/[h^{-1}M_{\odot}]) \geq 14$), is ~5%. It is important to note that as SZ observation sensitivity increases, observable-mass scaling relations will be measured using larger apertures, most notably $R_{ap} = 5R_{500c} \approx 2R_{200m}$. While our computation illustrates that with this larger aperture the assembly-driven scatter in $Y_{SZ}-M$ increases to 9%, it is possible that contributions to the intrinsic scatter from the cluster core (largely due to feedback) will be reduced. Hence, the optimal aperture radius that minimizes the intrinsic scatter in $Y_{SZ} - M$ is yet to be determined. Additionally, our model uses a very simple prescription for the gas density profile, with its only source of halo-to-halo variance, the mass-concentration relation, introducing less than 1% scatter into $M_{gas} - M$. Based on the NR simulation results of Stanek et al. (2010), we expect that by using a gas profile model that incorporates a more realistic response to halo assembly, scatter in $M_{gas} - M$ should increase (to ~3.6%) and covariance between M_{gas} and T_{mg} residuals should be significant (Pearson $\rho = 0.48$). Using these estimates, the scatter in our model $Y_{SZ} - M$ should increase to $7 - 12\%$ (in the range of $R_{ap} = R_{500c} - 5R_{500c}$), which is even closer to the results of simulation studies.

Observational studies tend to find a higher intrinsic scatter in the wider range of $14-35\%$, most of which use $R_{ap} = R_{500c}$, but similar results are found with $R_{ap} = R_{2500c}$. If a 5% (10%) Gaussian scatter is added to the cluster masses to mimic observational uncertainties, our predicted scatter in $Y_{SZ}(< R_{500c}) - M(< r_{500c})$ increases from ~5% to 10% (18%), which is more consistent with the observed results. The observational errors, particularly with regards to mass estimation, are still large; hence, the true intrinsic scatter in the relation is expected to be significantly lower than the values reported in the current observational literature, further motivating the development of more-accurate mass estimation techniques. However, it is also possible that additional processes not modeled in the full-physics simulations (e.g., magnetic fields and cosmic rays) are responsible for some of the additional intrinsic scatter observed.

5.3.4 Mass accretion rate prediction

As discussed in the Section 5.2.5, the model predicts a skewed distribution of the ln-residuals of the $Y_{SZ} - M$ relation due to the skewed distribution of MARs, Γ (see equation [5.12]). In the SK14 model, a high recent MAR will increase $f_{nth}(r)$, resulting in a decrease in the magnitude of the observables, T_{mg} and Y_{SZ} (at fixed halo mass). In Fig. 5.9, we plot the distributions of Γ and the ln-residuals, $\mathcal{R}$, computed for the $Y_{SZ}(< R_{200m}) - M(< r_{200m})$ relation at $z = 0$. There is a strong left-skew in the $\mathcal{R}$ distribution towards over-predictions, and this skewness is mirrored in the MAR distribution towards a small fraction of haloes with high Γ (i.e., disturbed clusters). The skewness in $\mathcal{R}$ is present regardless of mass cutoff or aperture employed.

The correspondence between the two distributions suggests that the $Y_{SZ} - M$ residual, which is itself an

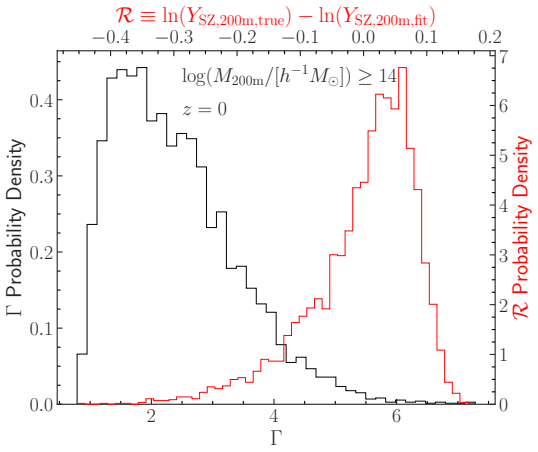

Figure 5.9: The distributions of halo MARs, Γ, defined by equation (5.12), and the ln-residuals, $\mathcal{R}$, of the $Y_{\rm SZ} - M$ relation. The scaling relation is computed at $z = 0$ for clusters in the mass range $14 \leq \log(M_{\rm 200m}/[h^{-1}M_\odot]) \leq 15.6$. The right-skew of the MAR distribution towards a minority of disturbed clusters is responsible for the left-skew in the $\mathcal{R}$ distribution, as a high MAR increases $f_{\rm nth}$ and reduces the magnitude of the observables. The correspondence between the distributions suggests an anti-correlation between the two quantities (see Fig. 5.10).

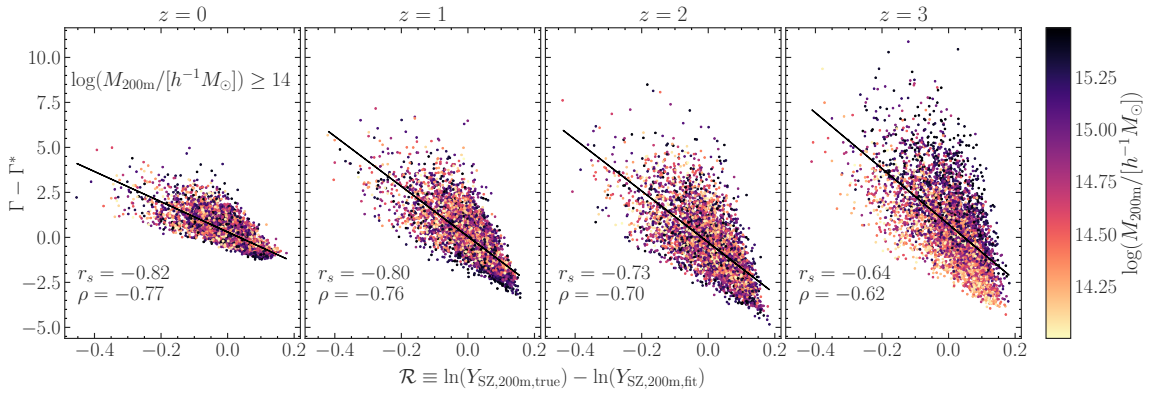

Figure 5.10: The relationship between the MARs and the $Y_{\rm SZ} - M$ ln-residuals for clusters at $0 \leq z \leq 3$ and in the mass range $\log(M_{\rm 200m}/[h^{-1}M_\odot]) \geq 14$. The MARs are plot relative to the median MAR of haloes of a given z and $\nu_{\rm 200m}$, denoted Γ^*, using the fitting function of Diemer et al. (2017). At higher Γ, a halo tends to have more non-thermal pressure, which reduces the magnitude of $Y_{\rm SZ}$, ultimately decreasing $\mathcal{R}$. The trend is strongest at $z = 0$, with a Pearson $\rho = -0.77$ and Spearman $r_s = -0.82$, weakening slightly as z increases. The slope in the relation tends to steepen with z. If present in full-physics simulations, this $\Gamma - \mathcal{R}$ relation may provide a link between the splashback radius and the *observable* residual, $\mathcal{R}$.

observable quantity, is likely to anti-correlate with the underlying halo MAR. The ability to estimate Γ from an observable would be powerful since, as discussed in Diemer et al. (2017), Γ is closely connected to the splashback radius and mass, $r_{\rm sp}/r_{\rm 200m}$ and $M_{\rm sp}/M_{\rm 200m}$. The splashback radius has been suggested as a better, physically-motivated definition for the halo boundary (Adhikari et al., 2014; Diemer & Kravtsov, 2014; More et al., 2015; Mansfield et al., 2017; Xhakaj et al., 2020), but it has proven difficult thus far to observe $r_{\rm sp}$ for individual clusters.

In Fig. 5.10, we plot the cluster MARs against their $Y_{\rm SZ} - M$ ln-residuals for several different redshifts. Diemer et al. (2017) present a fitting function for the median MAR seen in cosmological simulations as a function of z and $\nu_{\rm 200m}$, which we denote Γ^* and use to standardize our Γ values (i.e., $\Gamma - \Gamma^*$). There is a strong trend between the scaling relation ln-residual and the median-standardized MAR, with a Pearson $\rho = -0.77$ and Spearman $r_s = -0.82$ at $z = 0$. The slope of the relation tends to increase in magnitude slightly with z. Importantly, for $\mathcal{R} = 0$, the trend predicts that $\Gamma - \Gamma^* \approx 1$; in other words, if the cluster falls on the best-fit line for the $Y_{\rm SZ} - M$ relation, its MAR tends to be around the median for a halo of its mass at z. It is unclear whether or not such a strong trend between the residuals and halo MAR exists in real clusters, since previous simulation studies (e.g., Battaglia et al., 2012) have found that the MAR-driven skewness in the $Y_{\rm SZ} - M$ residuals decreases and the distribution approaches normality when additional physics beyond NR hydrodynamics is modeled in the simulations. The relationship between the observable $Y_{\rm SZ} - M$ ln-residuals and Γ should be explored in future full-physics simulation studies in order to quantitatively measure the strength of the $\Gamma - \mathcal{R}$ relation and forecast its predictive power for determining other secondary cluster properties that have recently been tied to the MAR, such as assembly bias (Sunayama & More, 2019) and the asphericity of the ICM (Chen et al., 2019).

5.4 Discussion

Our approach assumes that non-thermal pressure is dominated by turbulence generated during mass assembly. The SK14 model of the non-thermal pressure profile does not yet take into account various secondary effects due to baryonic physics, many of which will likely increase the intrinsic scatter in the scaling relations from what is presented here, especially for low-mass haloes and when small apertures are used. Radiative cooling and star formation results in the condensation of gas into the center of the cluster, reducing the baryon budget. Both Shaw et al. (2010) and Flender et al. (2017) modeled this by assuming that the gas adiabatically contracts or expands due to the change in total gas mass. Feedback due to AGN and supernovae provide additional sources of turbulence, especially in the inner regions of the cluster (e.g., Vazza et al., 2013; Zhuravleva et al., 2014; Chadayammuri et al., 2020). These feedback effects become more

significant as the halo mass decreases; hence, they must be accurately modeled in order to successfully quantify the thermodynamic properties of low-mass haloes through stacked measurements from upcoming microwave and X-ray surveys.

Observations of non-thermal X-ray emission and radio haloes (e.g., Million & Allen, 2009; Kale et al., 2013) imply the presence of additional non-thermal pressure due to cosmic rays and magnetic fields in the ICM. Substantial turbulent energy can also be injected into the cluster outskirts by the magneto-thermal instability (Parrish et al., 2008, 2012). Strong constraints have been placed on the magnetic field strength in the ICM, limiting the magnetic field-associated pressure to be much smaller than the thermal pressure ($\approx$ 5%, Dolag & Schindler, 2000; Iapichino & Brüggen, 2012). Observations of γ-ray emission in nearby clusters provide constraints on the pressure due to cosmic ray protons generated from shocks in the ICM to be less than 2% of the thermal pressure (Ackermann et al., 2014; Shirasaki et al., 2020a). Some simulations, however, suggest that cosmic rays could provide almost 50% of the total pressure in the cluster cores (Sijacki et al., 2008). Thus, although the overall additional non-thermal pressure due to magnetic fields and cosmic rays is likely small, better constraints are still warranted in order to determine the importance of incorporating their effects into future models of the total non-thermal pressure support.

Throughout our work, we assume spherically symmetric pressure and gas profiles for the clusters. The observational analysis of Arnaud et al. (2010) has shown that deviations from spherical symmetry and variations in cluster shapes can lead to scatter in the spherically-averaged pressure profiles. The recent hydrodynamical simulation study of Chen et al. (2019) has also reported that deviations from spherical symmetry increase the scatter in X-ray-based observable-mass scaling relations, additionally illustrating that the ellipticity of the ICM may be seeded by the MAH. Hence, the impact of mass assembly on the scatter in both SZ and X-ray observable-mass relations studied using our approach should still be regarded as a lower bound. The strength of future theoretical models will be greatly increased by incorporating the effect of mass accretion on triaxiality and cluster shape.

We have also neglected the impact of line-of-sight projection effects on the cluster observables. In particular, we study some spherically-integrated observables (i.e., $M_{\rm gas}$ and $T_{\rm mg}$) and cut off the $y_{\rm SZ}(R)$ line-of-sight integration at $2r_{\rm 200m}$. However, simulated light cone studies have demonstrated that a non-negligible fraction of the SZ signal arises from from the warm diffuse gas residing outside of groups and clusters (Hallman et al., 2007). Furthermore, Shirasaki et al. (2016) found that the projection of correlated structures along the line-of-sight introduces additional scatter into the scaling relation between the tSZ effect signal and the weak lensing mass. Thus, future gas models that aim to be combined with N-body simulations for efficient production of mock light cones must take into account the impact of the warm-hot intergalactic medium and other correlated structures along the line-of-sight.

5.5 Conclusion and future outlook

We quantified the effects of mass assembly-driven turbulence on the $Y_{SZ} - M$ scaling relation. This was accomplished by combining a simple model of the total pressure and gas density profiles, a model of the evolution of non-thermal pressure, and Monte Carlo-generated halo mass accretion histories. We summarize our most salient findings below:

- The average non-thermal pressure fraction profiles, $f_{nth}(r)$, tend to increase as halo mass or observation redshift increases. This is simply due to the fact that (i) higher mass haloes assemble later and (ii) a higher redshift of observation requires more rapid mass accretion at fixed halo mass.

- When radii are normalized by r_{200m}, the model predicts $f_{nth}(r/r_{200m})$ profiles that exhibit near-universality in redshift *at fixed peak height*, ν_{200m}. This finding is consistent with the simulation study of Nelson et al. (2014b). We provide a fitting function for $f_{nth}(r/r_{200m}|\nu_{200m})$ described by equation (5.21) and Table 5.3.

- As a consequence of $f_{nth}(r)$ increasing with halo mass and redshift, the model predicts that the magnitude of the average HSE mass bias (i.e., the deviation of the HSE-inferred mass from the true mass) also experiences these same trends.

- The scatter in the $Y_{SZ} - M$ relation due solely to inter-cluster variance in the halo MAHs ranges from $5 - 9\%$, increasing with aperture radius and z. This should be regarded as a lower bound, as the scatter will likely increase by a few percent once a more realistic model of the gas density profile is incorporated. For reference, most NR hydrodynamical simulations predict $10 - 15\%$ scatter. Thus, our model predicts that assembly-driven turbulence is responsible for a substantial fraction of the total scatter in $Y_{SZ} - M$.

- The slope of $Y_{SZ} - M$ tends to decrease slightly away from the self-similarity slope of $5/3$ as aperture, redshift, or halo mass limit increases. This dependence on aperture was also reported in Nagai (2006).

- There are small trends in the slope, normalization, and scatter of $Y_{SZ} - M$ with cosmology. The trends seen in Fig. 5.8 are consistent with those seen in the X-ray observable-mass relations of the simulated clusters in Singh et al. (2020). The perturbations in $Y_{SZ} - M$ due to cosmology are similar in magnitude to those seen due to variations in the lower mass cutoff used for computing the relation. This indicates that careful control of sample selection will be essential for any attempt to use cluster scaling relations to constrain cosmological parameters.

- The model predicts a skewed distribution of ln-residuals, $\mathcal{R}$, for $Y_{\rm SZ} - M$ due to the skewed distribution of Γ, in agreement with the NR hydrodynamical simulations of Battaglia et al. (2012). We find that Γ anti-correlates tightly with $\mathcal{R}$ (Spearman $r_s = -0.82$ at $z = 0$), potentially introducing a new observational approach to estimating the mass accretion rate via $Y_{\rm SZ} - M$.

The non-thermal pressure support present in galaxy clusters must be taken into account in order to make accurate HSE mass estimates and utilize the full statistical power that will be available in next-generation X-ray and SZ surveys for cluster count-based cosmological analyses. By studying the non-thermal pressure fraction profile, which is an important component of analytical models of the ICM (Shaw et al., 2010; Flender et al., 2017), we highlighted the dependence of accretion-driven turbulence on halo mass and redshift. As survey sensitivity continues to grow, the need to model and correct for the HSE mass bias over a wide range of halo masses and redshifts (especially smaller group mass haloes and high-redshift systems) is becoming increasingly important. This work represents a step towards developing a more accurate analytical model of the hot gas in groups and clusters, which will help (i) disentangle the effects of AGN/supernovae feedback from the non-thermal pressure driven by the structure formation process and (ii) model the cosmological dependence of the ICM.

The current model of the gas density, developed in KS01, is very simple and does not include any baryonic physics. Hence, a promising next step in model development should involve incorporating the effects of galaxy formation physics into the dark matter and gas density models (e.g., Schneider et al., 2020), which would enable the modeling of both galaxy formation and structure formation physics in a unified analytical framework. In addition, future cosmological simulations should focus on illuminating the importance of additional physical effects. Idealized simulations may over-predict the non-thermal pressure attributed to magnetic fields and thermal conduction (Parrish et al., 2012) since the turbulence and shocks generated by the structure formation process interact non-linearly with magnetic fields, which can lead to turbulence that changes non-monotonically with halo mass (McCourt et al., 2013). Modeling the turbulence pressure caused by additional sources as well as capturing baryonic effects on the dark matter (i.e., halo response modeling) and ultimately calibrating such models based on simulations will be crucial for combining the ICM model with models of the galaxy-halo connection and N-body simulations to generate a physically motivated and computationally efficient framework for interpreting forthcoming multi-wavelength cosmological datasets. Such an approach will eventually enable the use of correlation statistics from multi-wavelength cosmological surveys to constrain cosmology and astrophysics (Shirasaki et al., 2020b).

Currently, the best observations of bulk and turbulent motions in the ICM are of the Perseus cluster core, where the *Hitomi* X-ray observatory has reported high-resolution measurements of emission line Doppler

shifting and broadening (Hitomi Collaboration et al., 2018; Simionescu et al., 2019b, for a recent review). In the near future, *XRISM*/Resolve (Tashiro et al., 2018) and *Athena*/X-IFU (Barret et al., 2016) will measure the turbulence in the ICM for many nearby clusters and within the cores of more distant clusters, providing an opportunity to check the $f_{\rm nth}(r)$ model, correct for the HSE mass bias, and properly calibrate the mass scale (Ota et al., 2018). Furthermore, the *Lynx* X-ray Surveyor (Gaskin et al., 2019) and *Cosmic Web Explorer* (Simionescu et al., 2019a) have been proposed as future-generation X-ray telescopes that would enable exquisite measurements of turbulence out to the halo outskirts of an unprecedentedly large sample down to the galaxy mass scale.

In the future, millimetre-wave observations may provide a promising and complementary lens into the thermodynamics of and gas motions in the ICM via the thermal and kinematic SZ effects (see e.g., Mroczkowski et al., 2019). Upcoming and proposed microwave instruments, such as the TolTEC camera,[10] CCAT-prime,[11] CMB-HD (Sehgal et al., 2019), and Voyage2050 (Basu et al., 2021), will enable high-resolution SZ spectral imaging of clusters. This additional spectral information encodes a measurement of the kinematic SZ effect, which can be used to separate the cluster peculiar velocity and internal velocity dispersion (Inogamov & Sunyaev, 2003; Nagai et al., 2003; Sayers et al., 2019), thus providing a direct measurement of the non-thermal pressure support. Furthermore, since the strength of the SZ signal is independent of redshift, this approach can be used to observe the redshift-dependence of $f_{\rm nth}(r)$. Lastly, these observations will facilitate relativistic SZ corrections, which can be leveraged to study temperature structures in the ICM and mitigate the biases in the derived SZ and X-ray temperatures (see e.g., Chluba et al., 2012, 2013; Lee et al., 2020).

Finally, previous attempts at measuring the mass accretion rate of clusters via its relationship to the splashback radius have suffered from systematic uncertainties such as selection and projection effects (Baxter et al., 2017; Busch & White, 2017; Zu et al., 2017). The strong correlation between Γ and the ln-residuals, $\mathcal{R}$, of the $Y_{\rm SZ} - M$ relation highlighted in this study may provide an alternative means to measure the MAR, provided that the relationship is not washed out by other sources of non-thermal pressure or by observational errors. In addition to this $\Gamma - \mathcal{R}$ relation, machine learning algorithms may provide an alternative approach that enables more accurate determinations of both Γ and the cluster mass, employing input features such as images of the ICM and summary statistics that quantify the cluster shape (e.g., Green et al., 2019; Ntampaka et al., 2019a).

10. http://toltec.astro.umass.edu/

11. http://www.ccatobservatory.org/

Chapter 6

Cluster mass estimation

This chapter has been published as an article by Green et al. (2019) in the Astrophysical Journal by IOP Publishing.

6.1 Background

GALAXY clusters are the largest gravitationally bound objects in the universe. They are rare, with masses $\gtrsim 10^{14}\ M_{\odot}$, and their abundance is sensitive to the underlying cosmological model. Cluster counts can be used to constrain cosmological parameters, provided that there is an accurate way to connect the cluster observables (such as X-ray luminosity or temperature) to the underlying dark matter halo mass (for a recent review see Pratt et al., 2019).

Recent cluster-based constraints are in tension with *Planck* cosmic microwave background (CMB) cosmological constraints. For example, Sunyaev-Zeldovich (SZ; Sunyaev & Zeldovich, 1972) surveys find fewer massive clusters than would be expected from the *Planck* fiducial cosmology (e.g., Planck Collaboration et al., 2016b). This tension could be explained by a mass bias — a systematic under-estimation of X-ray based cluster mass estimates based on the hydrostatic assumption at the level of 30 – 45% (Planck Collaboration et al., 2016b; Bolliet et al., 2018; Zubeldia & Challinor, 2019; Makiya et al., 2020). However, a significant mass bias remains controversial. First, hydrodynamical cosmological simulations predict a hydrostatic mass bias in the range of 15 – 40% (e.g. Rasia et al., 2006; Nagai et al., 2007; Lau et al., 2013; Nelson et al., 2014a; Shi et al., 2016; Biffi et al., 2016; Henson et al., 2017) due to non-thermal pressure support provided by bulk and turbulent gas motions (e.g. Lau et al., 2009; Nelson et al., 2014b; Shi et al., 2015) and temperature inhomogeneities in the intracluster medium (ICM) (Rasia et al., 2014). Recent observational results agree that the hydrostatic bias must be small, at least for relaxed systems (e.g. Applegate et al., 2016; Eckert et al., 2019; Ettori

et al., 2019; Ghirardini et al., 2019). Second, the hydrostatic mass bias may also arise from the instrument-dependent systematic uncertainties in X-ray temperature measurements (Schellenberger et al., 2015; Israel et al., 2015). Finally, some cluster- and large-scale structure-based efforts put constraints on cosmological parameters that are consistent with those from the CMB (e.g., Mantz et al., 2015a; de Haan et al., 2016; Abbott et al., 2018) while others are in tension with them (e.g., Hildebrandt et al., 2020; Joudaki et al., 2020; Ntampaka et al., 2019b). Given the importance of this problem, concerted efforts are underway to calibrate the cluster mass scales using optical weak lensing measurements of background galaxies (e.g. von der Linden et al., 2014a; Hoekstra et al., 2015; Applegate et al., 2016; Dietrich et al., 2019) and CMB lensing (e.g., Raghunathan et al., 2019).

With next-generation observational surveys, such as the *eROSITA* X-ray survey (Merloni et al., 2012), soon to come online, massive data releases that will offer immense cosmological model constraining power are just around the corner. The *eROSITA* survey is predicted to identify ~93,000 galaxy clusters at or above the 50 photon limit with $M \gtrsim 10^{13.7}\, h^{-1}\, \mathrm{M}_{\odot}$ (Pillepich et al., 2012, 2018). The product of spectral temperature and gas mass, Y_X, is one of the lowest scatter mass proxies (Kravtsov et al., 2006). However, many of the *eROSITA* observations will be in the regime of low-photon counts, making T_X- and Y_X-based cluster mass estimates inaccessible (Borm et al., 2014). The core-excised luminosity ($L_{X,\mathrm{ex}}$) is another lower-scatter mass proxy that does not require T_X measurements; excluding the still poorly understood cluster cores ($r \lesssim 0.15\, R_{500c}$) reduces the scatter in the Y_X mass-$L_{X,\mathrm{ex}}$ (Maughan, 2007; Pratt et al., 2009) and weak lensing mass-$L_{X,\mathrm{ex}}$ (Mantz et al., 2018) relationships, but does so at the expense of drastically reducing the photon statistics.

Methods that provide improvements to L_X-based mass estimates for these low-photon *eROSITA* clusters could have a steep payoff. Even in the low-signal regime, there are subtle observable signals that can offer key insights for improving cluster mass estimates. Measures of cluster morphology, including surface brightness concentration (e.g., Santos et al., 2008), centroid shift (e.g., Mohr et al., 1993), and morphological composite parameters (e.g., Rasia et al., 2013), provide additional information about a cluster's dynamical state (Mantz et al., 2015b), which has been shown to influence the scatter in the mass-T_X relationship of simulated clusters (Ventimiglia et al., 2008), the correlated scatter in the relationship between weak lensing mass and integrated SZ Compton parameter Y_{sph} (e.g., Angulo et al., 2012; Marrone et al., 2012; Shirasaki et al., 2016), and the probability that a cluster is observed (Eckert et al., 2011; Planck Collaboration et al., 2011; Lovisari et al., 2017).

Modern machine learning (ML) techniques have been shown to reduce error in mass estimates of galaxy clusters. The techniques that have been developed use cluster dynamics (Ntampaka et al., 2015, 2016; Ho et al., 2019), X-ray images (Ntampaka et al., 2019a), and multiple wavelength summary statistics (Armitage

et al., 2019; Cohn & Battaglia, 2020) as input; similar ML techniques have also been applied to less-massive galaxy groups (Calderon & Berlind, 2019; Man et al., 2019). These methods hinge on using ML to extract additional information from complex correlations in the mass-observable relationships. Here, we use ML to take advantage of the complex correlations among morphological parameters, dynamical state, and cluster mass to improve mass estimates.

Our new X-ray cluster mass measurement technique utilizes cluster dynamical state information, encoded in X-ray morphological parameters, to provide improved, lower-scatter mass estimates relative to a mass-luminosity linear regression. In addition, we demonstrate that this improvement is obtained even in low-photon count *eROSITA* observations, which makes the inclusion of dynamical state information a promising avenue for future cosmological analyses that depend on robust cluster mass estimates. In Section 6.2, we introduce the *Magneticum* simulations and mock *Chandra* and *eROSITA* X-ray observations of simulated galaxy clusters used in this work. In Section 6.3, we provide an overview of the X-ray morphological parameters employed as features in our models. In Section 6.4, we describe the preprocessing of the mock catalog data and several regression methods used to build our models. We summarize the results of our models in Section 6.5, followed by our conclusions and proposed follow-up work in Section 6.6.

Throughout this chapter, the WMAP7 ΛCDM cosmology (Komatsu et al., 2011) is used: $\Omega_\mathrm{m} = 0.272$, $\Omega_\Lambda = 0.728$, $\Omega_\mathrm{b} = 0.046$, $h = 0.704$, $\sigma_8 = 0.809$, and $n_\mathrm{s} = 0.963$. The base-10 logarithm is denoted by log. All errors are quoted at the 68% level. The majority of this work is performed using the `scikit-learn` (Pedregosa et al., 2011) Python package.

6.2 Hydrodynamical simulations

6.2.1 The *Magneticum* simulations

Our cluster catalog is built from the *Magneticum*[1] (Dolag et al., 2015; Dolag et al., 2016; Ragagnin et al., 2017) suite of cosmological hydrodynamical simulations. *Magneticum* uses a WMAP7 cosmology (Komatsu et al., 2011) with a range of baryonic physics included (for additional details about the simulations and the included baryonic physics, see, e.g., Biffi et al., 2013; Steinborn et al., 2015; Teklu et al., 2015; Steinborn et al., 2016; Bocquet et al., 2016; Remus et al., 2017).

We select clusters from the *Magneticum* `Box2` and `Box2b` high-resolution simulations, selected for having sufficient resolution and volume to produce a suitable cluster catalog. `Box2` has cubic side length of $352h^{-1}$ Mpc with a dark matter particle resolution of $M_\mathrm{dm} = 6.9 \times 10^8\ h^{-1}\ \mathrm{M}_\odot$ and halo catalogs at $z = 0.10$,

1. https://www.magneticum.org/

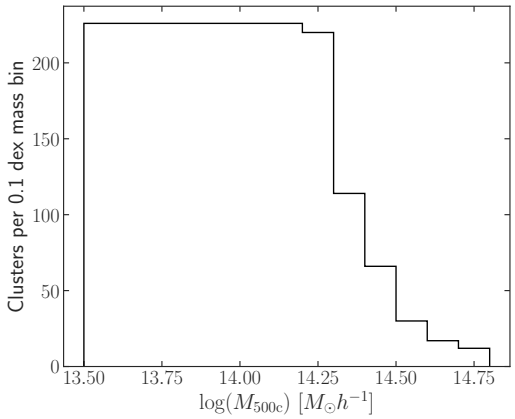

Figure 6.1: Mass function of the cluster sample used in this work. The sample is flat in the range $10^{13.5} \leq M_{500c}/(h^{-1}\,\mathrm{M}_\odot) \leq 10^{14.2}$ and begins decaying in cluster counts for $10^{14.2} \leq M_{500c}/(h^{-1}\,\mathrm{M}_\odot) \leq 10^{14.8}$. This sample consists of a total of 2,041 clusters. This uniform distribution in $\log(M_{500c})$, our predicted quantity, enables the regression model optimization to equally weight a broad range of cluster masses.

0.14, 0.17, 0.21, 0.25, and 0.29 (as well as higher z, but these are not included in our analysis). Box2b is larger in volume ($640h^{-1}$ Mpc on a side), has identical mass resolution, and has cluster catalogs at $z = 0.25$ and 0.29.

We initially select all clusters according to their spherical overdensity masses, M_{500c},[2] determined using the SUBFIND algorithm (Springel et al., 2001; Dolag et al., 2009). All clusters above $10^{13.5}\ h^{-1}\,\mathrm{M}_\odot$ are initially included and then subsampled in order to limit the sample to $\leq$ 230 clusters per 0.1 dex mass bin. The resulting training catalog has a flat mass function at lower masses, which helps to eliminate mass dependence in the scatter. Above $\sim 10^{14.2}\ h^{-1}\,\mathrm{M}_\odot$, the mass function of this sample falls off, following the mass function of the simulation (Bocquet et al., 2016). Hence, the sample has a flat mass function in the range $10^{13.5} \leq M_{500c}/(h^{-1}\,\mathrm{M}_\odot) \leq 10^{14.2}$ and a falling mass function in the range $10^{14.2} \leq M_{500c}/(h^{-1}\,\mathrm{M}_\odot) \leq 10^{14.8}$ (see Fig. 6.1). Redshifts in the range $0.1 \leq z \leq 0.21$ are roughly equally represented, with ~300 clusters per redshift. However, our sample contains ~450 clusters at $z = 0.25$ and $z = 0.29$ due to the addition of Box2b clusters.

The final cluster sample includes a total of 2,041 clusters in the redshift range $0.1 \leq z \leq 0.29$, consisting of clusters from both Box2 and Box2b. Within this sample, there are 984 unique clusters, many of which are observed at multiple redshifts. Based on the assumption of self-similarity (Kravtsov & Borgani, 2012), we verify that the distributions of all relevant features included in the model exhibit minimal redshift evolution, justifying our inclusion of multiple snapshots for a particular cluster.

2. We define M_{500c} as the mass enclosed within a sphere of (comoving) radius R_{500c} whose mean density is 500 times the critical density of the universe at $z = 0$.

6.2.2 Mock observations

From the cluster catalog, we create mock *Chandra* and *eROSITA* observations, employing the PHOX algorithm (Biffi et al., 2012, 2013). PHOX models the ICM thermal emission from gas particles by computing the expected number of photons given a fiducial (and large) exposure time and collecting area. The photon energies are then projected onto the sky plane along a chosen line of sight and cosmologically redshifted. A foreground galactic absorption model is applied, and models for *Chandra* ACIS-I and *eROSITA* are used to simulate the actual detections. Further details of the *Magneticum* implementation of PHOX can be found in Biffi et al. (2012), Biffi et al. (2013), and the publicly available *Magneticum* Cosmological Web Portal (Ragagnin et al., 2017).[3]

This implementation of PHOX allows the user to select from a number of parameters. For all observations, we select the ICM-only setting (i.e., AGN are not included as point sources in this work) and employ a 10 Mpc image line-of-sight size to include all relevant correlated structure. We seek to quantify the level of improvement in cluster mass estimates that can be made possible by incorporating dynamical state information, first in an idealized scenario and then in a realistic case that will be consistent with the observations made in upcoming large, high-throughput surveys such as *eROSITA*. To this end, our analysis features two different mock observation series: (i) mocks with *Chandra*-like angular resolution ("idealized *Chandra*" for short) with a *Chandra* ACIS-I instrument area and field of view (2071×2071 pixels, 16.9' FoV, 0.49" pixel) and a 1 Ms observing time, in the idealized regime of a flat effective area with respect to photon energy ($600\,\mathrm{cm}^2$) and no point spread function (PSF) smearing, as well as (ii) "realistic *eROSITA*" observations with an *eROSITA* instrument area and field of view (384×384 pixels, 1.03° FoV, 9.7" pixel) with a 2 ks observing time (Merloni et al., 2012) and instrument response and PSF modeled (see Ragagnin et al. (2017) for further details regarding *eROSITA* instrument modeling). To more closely imitate the conditions of the upcoming *eROSITA* observations, the "realistic *eROSITA*" mock images also include background noise. The process by which this noise is added is described below. The *eROSITA* mock observations have a median photon count of ~2000 for clusters observed at $z = 0.1$ and ~100 for clusters observed at $z = 0.29$. In contrast, the *Chandra* mocks have a median photon count of $\sim 6 \times 10^5$ for clusters observed at $z = 0.1$ and $\sim 3 \times 10^4$ for clusters observed at $z = 0.29$; clearly, derivative quantities computed from the "idealized *Chandra*" observations will be affected much less by Poisson noise.

The cluster bolometric luminosities L_X are calculated by PHOX using the publicly available X-ray package XSPEC (Arnaud, 1996). Core-excised luminosities $L_{X,\mathrm{ex}}$ are computed as follows: (i) compute the total observed photon count N_{tot} within R_{500c}, (ii) compute the observed photon count within $0.15R_{500c}$, denoted

3. https://c2papcosmosim.uc.lrz.de/

N_{ce}, and (iii) scale the bolometric luminosity by the ratio of the photon count observed outside of the core to the total photon count, i.e., $L_{X,ex} = \frac{N_{tot}-N_{ce}}{N_{tot}} L_X$. In this work, the core-excised luminosity is used since it has been shown to have lower intrinsic scatter with the cluster mass (Maughan, 2007; Mantz et al., 2018) and is less sensitive to the details of the complicated core physics models used in the simulations. We note that, for the "realistic *eROSITA*" observations, the core-excised photon count ratios are computed *prior to the addition of background noise*. This likely makes our mock luminosities more accurate than in the case of real *eROSITA* observations. However, this choice puts the core-excised luminosities from our two observation series on equal footing, such that the model performance differences between the idealized and realistic cases will be dominated by the quality of the morphological parameters.

Redshift is not explicitly included as a feature to train the regression models, however the redshift is used in scaling the luminosity. Thus, the core-excised luminosity used in this work is always appropriately scaled by the redshift evolution factor, assuming self-similarity, such that we use

$$L_{ex,z} \equiv L_{X,ex} E(z)^{-7/3} = \frac{N_{tot} - N_{ce}}{N_{tot}} L_X E(z)^{-7/3}. \tag{6.1}$$

In the $0.5 - 2.0\,\mathrm{keV}$ energy band, *eROSITA* anticipates an average photon plus particle background of 2.19×10^{-3} counts s^{-1} arcmin^{-2} (Clerc et al., 2018). Thus, for *eROSITA*, the background is given by a Poisson distribution with rate $\lambda = 0.113/(2\,\mathrm{ks})$ (Merloni et al., 2012). A unique Poisson background is generated for and added to each *eROSITA* mock observation.

6.3 Morphological parameters

In order to encode information about the dynamical state of the cluster into the model, we incorporate various morphological parameters as features, all of which can be directly calculated from the mock X-ray images. In the following, we define each of these parameters. We refer the reader to Lotz et al. (2004), Rasia et al. (2013), and Lovisari et al. (2017) for more in-depth discussion on each of the parameters. Unless otherwise specified, the aperture used to compute the morphological parameters has a radius of $R_{ap} = R_{500c}$ and is centered on the cluster X-ray peak; we discuss the implications for this choice at the end of this section.

First, the concentration parameter c quantifies how centrally concentrated the X-ray emission is within the cluster, and has been shown by Santos et al. (2008) to be an indicator for the presence of cooling-core systems at high z. Concentration is defined to be the ratio of the flux within two circular apertures: $0.1R_{ap}$ and R_{ap} (Lovisari et al., 2017).

The centroid shift parameter w is defined as the variance of the projected separation between the X-ray peak of the image and the emission centroid obtained within 10 circular apertures of increasing radius up to $R_{\rm ap}$ (Lovisari et al., 2017).

The power ratios, introduced by Buote & Tsai (1995), use the ansatz that the X-ray surface brightness profiles are a good tracer of the cluster's projected mass distribution. The "power" is encoded in the coefficients of a 2D multiple decomposition of the cluster X-ray image, where higher-order components probe increasingly smaller scales. The nth-order power ratio $P_{n0} = P_n/P_0$ is, in essence, the ratio between the nth multipole moments and the 0th multipole moment. In this work, we consider P_{10}, P_{20}, P_{30}, and P_{40}. The latter two probe large- and small-scale substructures present within the cluster, and thus further convey dynamical information.

The second power ratio P_{20} provides a measurement of the cluster ellipticity. Another ellipticity parameter, denoted e, is also calculated, defined as the ratio between the semiminor and semimajor axis (Lovisari et al., 2017).

The asymmetry parameter A quantifies the rotational symmetry of the cluster X-ray emission (Lotz et al., 2004). A is calculated by rotating by 180° and self-subtracting the background-subtracted cluster image from itself, summing the values of the pixels in this image difference and normalizing by the summed pixels in the original image (Abraham et al., 1996).

The smoothness S quantifies the degree of small-scale substructure within the cluster (Lotz et al., 2004). S is calculated by boxcar-smoothing and self-subtracting the background-subtracted cluster image from itself, again summing the values of the pixels in this image difference and normalizing by the summed pixels in the original image (Conselice, 2003).

Lastly, the M_{20} parameter is an analog of concentration (Lotz et al., 2004). The total second-order moment of the light is a distance-to-center-weighted sum of the flux f_i within all pixels i in the cluster, $M = \sum_i f_i[(x_i - x_{cc})^2 + (y_i - y_{cc})^2]$, where cc denotes the cluster center. Then, M_{20} is computed as the ratio of the partial second moment M_p, which sums over only the brightest pixels that contain 20% of the cluster light, divided by the total second moment, written as $M_{20} = \log(M_p/M)$.

These morphological parameters encode dynamical state information. For example, disturbed clusters tend to be asymmetric (high A), clumpy (*high* S), and not concentrated (low c). All of the parameters introduced above are calculated for each mock cluster observation, and are used as features in our regression model.

In the subsequent analysis, we utilize two distinct series of morphological parameters, which are computed from our two mock observations series, described above. The "idealized *Chandra*" series is computed from the background-free *Chandra* observations, using $R_{\rm ap} = R_{\rm 500c}$. The "realistic *eROSITA*" series is com-

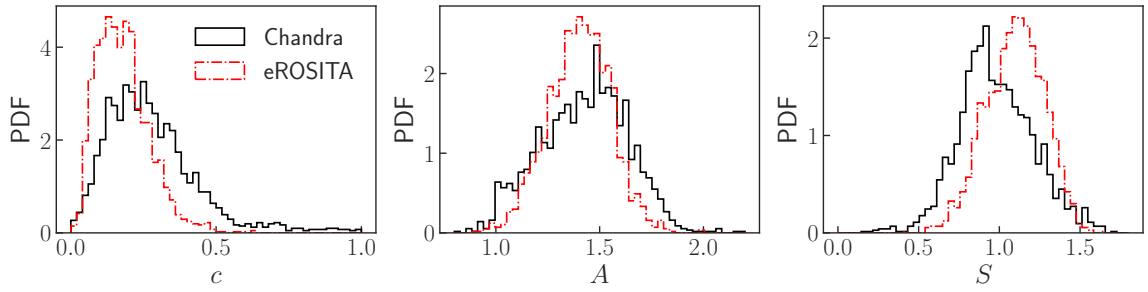

Figure 6.2: Distributions of the surface brightness concentration c, asymmetry A, and smoothness S computed from the "idealized *Chandra*" and "realistic *eROSITA*" mock cluster observation series. Several other morphological parameters are also employed in the analysis (see Sec. 6.3), but we find that c, A, and S are most important for strengthening the cluster mass model.

puted from the *eROSITA* observations with added background, also using $R_{\rm ap} = R_{\rm 500c}$; in this case, the mean background is subtracted prior to computing the parameters. The former series is intended to give an upper bound on the expected improvement in cluster mass estimates made possible by including dynamical state information present in idealized, high spatial resolution (0.5"), and high-photon count observations. The latter series is intended to give a more realistic estimate of the expected improvement that will be possible in upcoming cosmological analyses that will be performed with low-photon count cluster observations. We acknowledge that by using the exact $R_{\rm 500c}$ for our aperture when computing the morphological parameters, we are neglecting additional scatter that will be present due to this effect. Also, in the case of the "idealized *Chandra*" series, properly including PSF effects would introduce additional smoothing to these observations. Hence, our subsequent results will remain as optimistic estimates. As we find below, the most important morphological parameters are smoothness, asymmetry, and concentration. In Fig. 6.2, we plot the distributions of these three parameters, comparing the "idealized *Chandra*" and "realistic *eROSITA*" series. While we find generally good agreement between the two series, it is clear that the *eROSITA* cluster observations result in systematically lower concentrations and higher smoothness parameters.

The lack of high-concentration objects in the *eROSITA* mocks is due to the broader PSF of *eROSITA* with respect to *Chandra*. Photons originating from the central regions of the observed systems are redistributed over a wider area, which reduces the concentration with respect to the true value. Since we do not attempt to correct for PSF smearing by applying PSF deconvolution, our procedure for reconstructing c values from *eROSITA* mocks underestimates the concentration of highly-peaked objects. While also impacted by the broader PSF, the shift to larger S (i.e., *less* smooth) in the *eROSITA* mocks is additionally due to both (i) the lower exposure time, which results in a less "filled in" photon distribution due to Poisson noise, and (ii) the presence and subtraction of background, which introduces additional Poisson noise.

Additionally, in Fig. 6.3, we show several example *Chandra* cluster images to demonstrate the morpho-

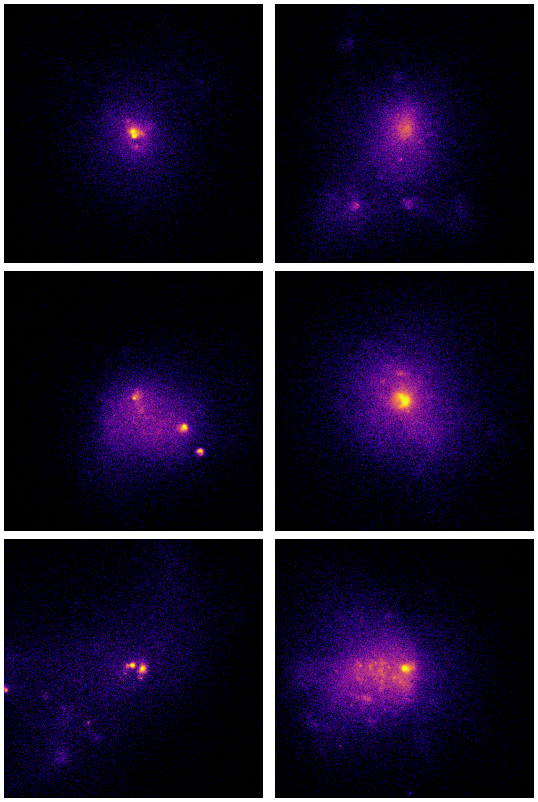

Figure 6.3: Sample *Chandra*-like cluster images illustrating morphological parameter differences. Each image is centered on the cluster X-ray peak and is cropped a side length of $2R_{500c}$. All clusters shown have $10^{14.3} \leq M_{500c}/(h^{-1}\,\mathrm{M}_{\odot}) \leq 10^{14.6}$ at $z = 0.1$. Top: highly concentrated cluster on the left ($c = 0.37$) and weakly concentrated cluster on the right ($c = 0.04$). Middle: Asymmetric cluster on the left ($A = 1.49$) and symmetric cluster on the right ($A = 0.93$). Bottom: Less smooth cluster on the left ($S = 1.03$) and more smooth cluster on the right ($S = 0.61$). Note that a higher value of the smoothness parameter corresponds to a cluster whose surface brightness profile is less smooth. When combined, these parameters, among others (see Sec. 6.3), capture the cluster dynamical state by quantifying details such as the presence of substructure or tidal distortions.

logical parameters, in particular the concentration c, asymmetry A, and smoothness S. All clusters shown have roughly the same mass, lying in the range $10^{14.3} \leq M_{500c}/(h^{-1}\,\mathrm{M}_\odot) \leq 10^{14.6}$, and are all at the lowest redshift of $z = 0.1$. The images are all scaled by the cluster R_{500c}. Clearly, a cluster with a larger concentration has a substantially larger fraction of its flux coming from its core. Furthermore, the asymmetry parameter is successfully able to capture disturbances or substructure in the cluster that result in reduced symmetry. The smoothness parameter is capable of quantifying small-scale structures; note that a cluster with a larger value of S is overall less smooth, and more likely to contain substructures.

6.4 Analysis methods

6.4.1 Data preprocessing

As stated previously, our sample consists of 2,041 mock cluster observations across six redshifts in the range $0.1 \leq z \leq 0.29$. Each observation consists of many features, including core-excised $L_{\mathrm{ex},z}$ and all of the morphological parameters described in the previous section. The logarithm of the power ratios P_{i0}, centroid shift w, and luminosity are used due to their large dynamic ranges, whereas the remaining features are not transformed. The regression target for each observation is $\log(M_{500c})$ of the cluster.

The sample is then split into a training set that comprises 80% of the observations (1,617 clusters) and a test set that comprises the remaining 20% of the observations (425 clusters); this train-test split is a common rule of thumb based on the Pareto principle. The split is performed such that all redshift observations of each unique cluster are assigned either to the training or test set, but not split between the two. For optimization of hyperparameters, k-fold cross-validation is employed on the training set, with $k = 10$. The folds are generated such that all observations of each unique cluster are confined to only one fold.

Many regression algorithms require the distribution of each observable to be scaled to have roughly zero mean and unit variance. In order to scale in such a way that is robust to outliers, we subtract the median and divide by the 1σ (16th/84th) percentile range computed over the *training* set in order to standardize each feature. The medians and σ are stored from the training set such that an identical transformation is applied to the test set.

6.4.2 Regression methods

The work of Armitage et al. (2019) found that ordinary linear regression (OLR) and ridge regression (RR; Hoerl & Kennard 1970) models were able to produce the least scatter in cluster mass estimates using a variety of X-ray, spectroscopic, and photometric datasets. The authors also tested an ordinary decision tree model

(Quinlan, 1986), as well as AdaBoost (Freund & Schapire, 1996) and gradient boosted regression (Friedman, 2001), but did not test the popular random forests (RF) regression algorithm (Breiman, 2001). Motivated by their work, and considering that our feature set of morphological parameters contains different information, we will focus our analysis on various linear regression methods and expand by applying random forest regressors. We train different regression models on our mock catalogs, including a standard mass-luminosity power law ($M - L_{\rm ex,z}$), an OLR model, several regularized linear regression models (including RR and Lasso regression [LR; Tibshirani 1996]), and RF regression models.

Ordinary linear regression is performed as follows. For n clusters, each of which are described by p features (i.e., observables), one has a data matrix $\mathbf{X} = \{\mathbf{x}_0^T, \mathbf{x}_1^T, \mathbf{x}_2^T, ..., \mathbf{x}_n^T\}$, where each $\mathbf{x}$ is a vector of length p. Each vector of observables $\mathbf{x}_i$ is associated with a true logarithmic mass y_i. The mass is predicted as a linear function of the observables, $y_i = \mathbf{x}_i^T \mathbf{m}\beta + \epsilon_i$, where $\mathbf{m}\beta$ is the model parameter vector of length p and ϵ_i is the random error in the model for cluster i. The best-fit model parameters are chosen by minimizing the cost function, which is selected to be the sum of squared residuals, $E_{\rm OLR}(\mathbf{m}\beta) = \sum_{i=1}^{n}(y_i - \mathbf{x}_i^T \mathbf{m}\beta)^2$.

In this work, we consider several OLR models. First, we train a simple OLR model with only one feature, the core-excised luminosity $L_{\rm ex,z}$. This allows us to set a baseline for performance and compare the results of our mock observations in terms of the scatter to an observed mass-luminosity relationship. Then, we train an OLR model on the full feature set, including core-excised luminosity and all morphological parameters.

In an effort to reduce the feature dimensionality and highlight the most important features in the model, one can use regularized linear regression, where an additional term is added to the cost function that introduces a penalty for models with large $\|\mathbf{m}\beta\|$. In ridge regression, the new cost function is of the form $E_{\rm RR}(\mathbf{m}\beta;\alpha) = E_{\rm OLR}(\mathbf{m}\beta) + \frac{\alpha}{2}\|\mathbf{m}\beta\|_2^2$, where $\|\cdot\|_2$ denotes the Euclidean norm. Similarly, in Lasso regression, the cost function is instead $E_{\rm LR}(\mathbf{m}\beta;\alpha) = E_{\rm OLR}(\mathbf{m}\beta) + \alpha\|\mathbf{m}\beta\|_1$, where $\|\cdot\|_1$ denotes the Manhattan norm.

Regularization acts to reduce the weights of unimportant features in the model, reducing the capability of the model to overfit the training data. Lasso regression is a more strictly regularized model than ridge regression. The hyperparameter α is selected via a grid-search cross-validation (CV) of logarithmically-spaced α values, where the model performance is evaluated via k-fold CV for each α. In k-fold CV, the training set is split into k random subsets (split according to unique cluster ID; see Sec. 6.4.1). Then, $k-1$ of the subsets are used to train the model, and model predictions are made on the remaining subset. This process is iterated k times such that predictions are made for all clusters in the training set. The model performance is quantified by the mean squared error (MSE) of all of the predictions. The CV process is repeated for all α in the grid, and the α that minimizes the MSE is selected for the training of the final model, which is trained on the full training set and subsequently applied to the test set.

For our last set of models, we use the non-parametric random forest regression model, which is an en-

semble technique based on decision trees. RF models reduce the issues of overfitting that are endemic to decision trees by randomly growing an ensemble of trees, each trained on a different subset of the total training data, and taking the average of their predictions. Furthermore, RFs increase the tree diversity relative to a standard decision tree ensemble by splitting each node according to the best feature in a random subset of the features, instead of the full feature set. This increased tree diversity results in a more generalizable model that is less prone to overfitting the training set. RFs have several important hyperparameters: (i) the number of trees in the forest, (ii) the maximum number of features that can be included in one node splitting condition, (iii) the maximum depth allowed for a tree (i.e., number of decisions that must be made to reach an output), (iv) the minimum number of samples in the training set at a particular node that are required in order for the node to split, (v) the minimum number of samples in the training set required to form a leaf, and (vi) whether or not to use bootstrap resampling (i.e., using "bagging" vs. "pasting"). Reducing the "maximum" hyperparameters (i.e., [ii] and [iii]) or increasing the "minimum" hyperparameters (i.e., [iv] and [v]) is an effective way to regularize the model and reduce the tendency for overfitting. The interested reader should refer to Géron (2017) for additional details of various machine learning regressors, including ensemble and tree-based regression.

In this work, we consider several RF models with different sets of hyperparameters and different input feature sets in order to demonstrate the level of sensitivity that RF models have to the hyperparameters and to tune an optimal model for future mass predictions. The first is a RF model with the default hyperparameters from the SCIKIT-LEARN implementation. The hyperparameters of the second model are optimized using grid search CV over the six-dimensional parameter space of hyperparameters described above. The third model includes a reduced set of features ($L_{\rm ex,z}$, S, A, and c), but the hyperparameters are also tuned via grid search CV.

After selecting hyperparameters for the various models using CV on the training set, the final models are each trained on the entire training set. The models are then applied to predict the masses of the test set, which we emphasize was never used for either hyperparameter selection or model training, and thus should represent a true example of the generalization capability of the models. The entire preprocess-split-cross-validate-train-test procedure is performed separately for each of the two series of morphological parameters, i.e., those from the "idealized *Chandra*" and "realistic *eROSITA*" observations. In the next section, we report the results for these final models as applied to the test sets.

	Chandra		*eROSITA*	
	Correlation r with		Correlation r with	
Feature	$\log(M_{500c})$	$\mathcal{R}$	$\log(M_{500c})$	$\mathcal{R}$
$\log(L_{ex,z})$	0.927	0.000	0.929	0.000
c	−0.101	−0.041	0.208	−0.146
e	0.060	−0.105	0.071	−0.111
$\log(w)$	−0.096	0.195	−0.126	0.228
$\log(P_{10})$	−0.170	0.212	−0.197	0.224
$\log(P_{20})$	−0.140	0.155	−0.308	0.194
$\log(P_{30})$	−0.139	0.146	−0.427	0.198
$\log(P_{40})$	−0.145	0.149	−0.489	0.207
A	−0.294	0.034	−0.654	0.131
S	−0.493	0.048	−0.795	0.119
M_{20}	0.032	0.116	−0.148	0.148

Table 6.1: Pearson correlation between each observable in the model and (i) the true mass, $\log(M_{500c})$, or (ii) the logarithmic mass residual from a mass-luminosity regression, $\mathcal{R}$. These calculations were performed using both the "idealized *Chandra*" and "realistic *eROSITA*" series of morphological parameters. In both series, S and A correlate most strongly with $\log(M_{500c})$. The correlations with mass are generally stronger in the "realistic *eROSITA*" series. While $\log(w)$ and $\log(P_{10})$ correlate most strongly with $\mathcal{R}$ in both series, we find that they are not the most important morphological parameters (rather, S, A, and c are).

6.5 Results

For both series of observations, we compute the Pearson correlation coefficient between each observable and the cluster mass, shown in Table 6.1. Additionally, the best fit linear regression model between $\log(M_{500c})$ and $\log(L_{ex,z})$ is used to make mass predictions, and the corresponding mass residuals are then correlated against the observables, also shown in Table 6.1. The mass residuals $\mathcal{R}$ are defined as

$$\mathcal{R} = \log(M_{500c,pred}) - \log(M_{500c,true}), \tag{6.2}$$

where we again emphasize that the base-10 logarithm is used throughout.

After luminosity, the observables that correlate or anti-correlate most strongly with mass are smoothness S and asymmetry A. The centroid shift w and first power ratio P_{10} correlate most strongly with the mass residuals, although these correlations are still quite weak ($|r| \approx 0.2$). Thus, the naive expectation is that w and P_{10} should be the most important additional features (i.e., after luminosity) in a multivariable model of the mass. However, as we will show below, this ends up not being the case. We note that while the ranking of the morphological parameters in terms of their correlation strengths remains close to the same between the two mock observation series, the strengths are systematically stronger in the "realistic *eROSITA*" observations. In particular, the high-order power ratios, smoothness, and asymmetry (i.e., the parameters that quantify substructure) correlate much more strongly with mass in the *eROSITA* observations, which is likely a result of the deviation from a smooth profile driven by Poisson noise in the low-photon

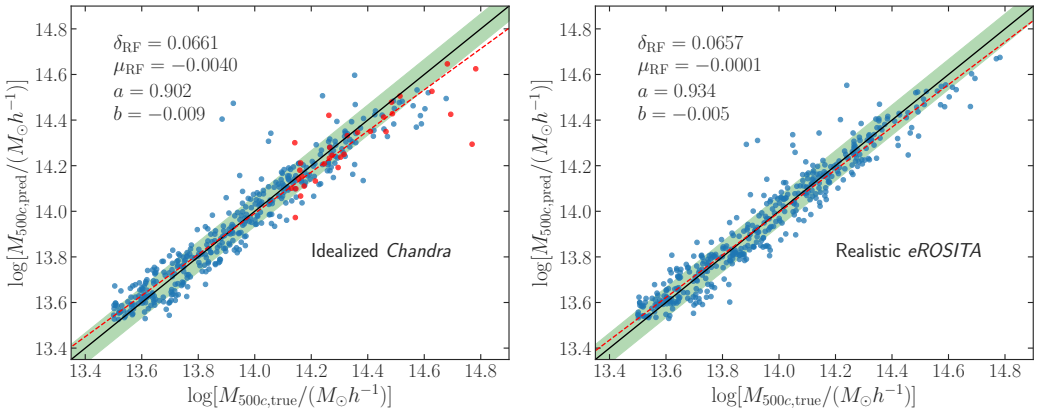

Figure 6.4: Predicted mass as a function of true mass. Predictions are made using the cross-validation-tuned RF models, which are separately trained using each of the two morphological parameter series. The distributions both have low intrinsic scatter δ (0.066 dex) and a negligible bias μ. The green band corresponds to the 1σ scatter, δ. The dashed red line shows the best power law fit to the predicted masses, $\log(M_{\rm 500c,pred}/[10^{14}h^{-1}\,{\rm M}_\odot]) = a\log(M_{\rm 500c,true}/[10^{14}h^{-1}\,{\rm M}_\odot]) + b$. For true masses above $\sim 10^{14.4}\,h^{-1}\,{\rm M}_\odot$, the model consistently underpredicts the mass. This is due to the falling mass function of our sample in the high-mass regime. Additionally, in the case of the *Chandra* observations, some clusters (indicated as red points) extend beyond the instrument field of view, which likely contributes to the lower accuracy of their predicted masses. Using a training set with a flat mass function that covers the full cluster mass range of interest will likely ameliorate these underpredictions, resulting in a predicted-to-true slope A closer to unity.

count regime. On the other hand, the correlations between the *Chandra* morphological parameters and the cluster masses are in good qualitative agreement with Lovisari et al. (2017), which, using *XMM-Newton* cluster observations, found no significant correlation between the total mass and c, w, or the power ratios.

The primary model of interest is our cross-validated random forest regressor, which, as we will show below, performs the best among all of the regression methods tested for both series of morphological parameters. The mass predictions generated by the random forest model for the 426 clusters in the test set are shown in Figure 6.4, with the two separate panels corresponding to the models trained and tested on the two different series of mock observations. In both cases, it is clear that the model begins to systematically underpredict the masses of the high-mass clusters with $M_{\rm 500c} \gtrsim 10^{14.4}\,h^{-1}\,{\rm M}_\odot$, which roughly corresponds to the regime where our sample transitions from a flat to falling mass function. In order to employ this method to predict the masses of observed clusters, it is crucial that the training sample consists of a flat mass function that covers the entire range of masses of interest. The performance of machine learning models, such as RFs, will greatly improve as larger training samples that are uniform in the prediction (in this case, the mass) become available, for example from state-of-the-art cosmo-hydrodynamical simulations.

The PDFs of the mass residuals for these cross-validated random forest regression models are shown in Figure 6.5. Additionally, the 1σ intrinsic scatter (i.e., half of the 16th–84th percentile range of $\mathcal{R}$) in the

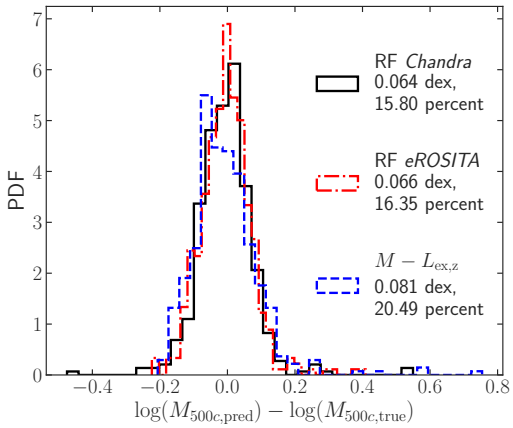

Figure 6.5: PDF of mass residuals for the cross-validation-tuned RF models of both the "idealized *Chandra*" and "realistic *eROSITA*" observation series. For comparison, we plot the PDF of mass residuals for the mass-luminosity relationship, with $L_{\mathrm{ex,z}}$ computed using core-excised luminosities from the "idealized *Chandra*" observations. In both cases, the RF model offers a ∼20% reduction in scatter relative to the mass-luminosity approach, with negligible bias.

test set for each of the trained models and both of the observation series are shown in Table 6.2. For our test sample, the mass residuals of the standard mass-luminosity relationship have a bias of $\mu = -0.017$ dex and 1σ scatter of $\delta = 0.081$ dex. Interestingly, for both the "idealized *Chandra*" and "realistic *eROSITA*" observations, the mass residuals have virtually negligible biases and 1σ intrinsic scatter of $\delta = 0.066$ dex, which amounts to a 20% reduction in scatter relative to the mass-luminosity relationship. Table 6.2 demonstrates that ordinary linear regression with a combined input feature set that includes the luminosity and all morphological parameters improves only marginally over the single variable mass-luminosity regression. The incorporation of regularization (i.e., the RR and LR models) does not result in an improved model. The lack of improvement in these linear models after the inclusion of morphological parameters illustrates that the relationship between cluster morphology and mass is nonlinear and justifies the use of nonlinear approaches, such as a RF regressor.

In the "idealized *Chandra*" observations, 140 of the most massive clusters have R_{500c} that extend beyond the instrument field of view. This results in changes to the morphological parameters calculated for these clusters; for example, the concentrations will systematically increase (although only slightly since the cluster outskirts have the lowest surface brightness) and parameters that quantify substructure (S, P_{30}, P_{40}) may deviate from the correct value if substructures lie outside of the field of view. We verified that the reported scatters are insensitive to the presence or removal of these clusters from the dataset. However, this effect, in addition to the dearth of high-mass clusters in the training sample, is likely responsible for the less accurate mass predictions for high-mass clusters (and lower predicted-to-true slope a) when using the *Chandra* observation series.

Method	$\delta_{\rm iC}$	$\delta_{\rm re}$
$M - L_{\rm ex,z}$	0.081	0.081
OLR	0.078	0.080
RR	0.078	0.080
LR	0.079	0.081
RF, defaults	0.070	0.067
RF with CV	0.066	0.066
RF, only $L_{\rm ex,z}$, c, A, S	0.070	0.071

Table 6.2: The 1σ percentile intrinsic scatter, defined as half of the 16th–84th percentile range of the mass residuals $\mathcal{R}$, for each model as trained in the text when applied to the test set, computed using the "idealized *Chandra*" and "realistic *eROSITA*" mock observation series, denoted $\delta_{\rm iC}$ and $\delta_{\rm re}$, respectively. The multivariable linear models (OLR, RR, LR) improve only marginally relative to the mass-luminosity relation, with regularization yielding no improvement. The RF models, which capture nonlinear relationships between the input features and the mass, are able to further reduce the scatter beyond any linear approach.

The intrinsic scatter for core-excised luminosity-based mass estimates in the observational literature ranges from $\lesssim$ 15% for the weak lensing mass-$L_{X,\rm ex}$ relationship (Mantz et al., 2018) to 16–21% for the Y_X mass-L_X relationship (Maughan, 2007). While the work of Mantz et al. (2018) finds a lower scatter than we do for our $M - L_{\rm ex,z}$ relation, we note that they employ a mass cutoff of $M \geq 3 \times 10^{14}\ M_\odot$, which results in a much smaller mass range than covered by our dataset. The scatter in the $M - L_{\rm ex,z}$ mass residuals for the subset of our clusters with $M_{\rm 500c} \geq 3 \times 10^{14}\ M_\odot$ is 17%, which is rather close to that of Mantz et al. (2018). The sample used by Maughan (2007) includes clusters down to $8 \times 10^{13}\ M_\odot$, which is more consistent with our cluster sample. Thus, the scatter in the $M - L_{\rm ex,z}$ mass residuals of our full sample, $\delta = 0.081$ dex (20.5% scatter), is consistent with similar such calculations performed using observations of either the Y_X mass-$L_{X,\rm ex}$ or weak lensing mass-$L_{X,\rm ex}$ relationships. The current state-of-the-art mass estimation methods require high-resolution, long-exposure cluster observations with good spatial and spectral resolution, with the Y_X approach resulting in 5–7% scatter (Kravtsov et al., 2006). The observational conditions necessary for utilizing the $M - Y_X$ method will simply not be present for the vast majority of clusters observed in upcoming surveys such as *eROSITA*. However, we have demonstrated that our method achieves a 20% improvement in cluster mass estimates over $M - L_{\rm ex,z}$ even in the low-spatial resolution, short-exposure (2 ks) conditions of *eROSITA* observations ($\delta = 0.066$ dex, 16%). Since we also find the same level of improvement for our *Chandra* observations, which likely places an upper bound on the method performance, this suggests that the mass-encoding dynamical state information, as quantified by our set of morphological parameters, remains present even in the short-exposure *eROSITA* observations.

One metric available for interpreting the results of the random forest model is the *feature importance* ranking. For example, in an OLR model, the feature importances are roughly quantified by the magnitudes of the regression coefficients. The standard metric for RF feature importance, and the one that is implemented in SCIKIT-LEARN, is the *mean decrease in impurity*, which measures how effective a feature is at reducing the

	Importance (%)	
Feature	*Chandra*	*eROSITA*
$L_{\mathrm{ex,z}}$	74.9	76.2
S	10.0	14.6
A	4.8	2.5
c	3.9	1.5

Table 6.3: The feature importances for the cross-validated RF models, computed based on the mean decrease in impurity. The remaining morphological parameters are omitted from the table, as their importances are all $\lesssim$ 1%. The smoothness S, asymmetry A, and surface brightness concentration c encode the most additional mass information after the core-excised luminosity $L_{\mathrm{ex,z}}$.

variance of the predictions. Based on this importance measurement, the most important features and their *Chandra* importances are (after $L_{\mathrm{ex,z}}$, 75%), in decreasing order: smoothness S (10%), asymmetry A (5%), and concentration c (4%). The ranking of these features are the same for both observation series, with importance magnitudes being similar, and all other morphological parameters have negligible importance ($\lesssim$ 1%) in both cases (see Table 6.3).

Another measurement of feature importance, known as *permutation importance*, quantifies the drop in the R^2 score when the values of a feature are permuted over the samples. Thus, the larger the drop in R^2 when a feature is permuted, the more important the feature. This metric is considered less biased, as it is not sensitive to the dynamic range of the input variables; this detail is likely irrelevant since our features are scaled. When this importance metric is employed, we find the same feature importance ranking as before for the "realistic *eROSITA*" parameter series. However, for the "idealized *Chandra*" series, we find that, following luminosity, c and S are nearly tied for being most important, followed by A. These three features are not the highest correlators with mass residual, which goes against the naive expectation that the most important features after luminosity should correlate the strongest with mass residual. However, S and A do correlate the most strongly with mass after $L_{\mathrm{ex,z}}$, as seen in Table 6.1. Thus, there must be some nonlinear relationship between $L_{\mathrm{ex,z}}$, S, A, and c (slightly supplemented by the combination of all the other morphological parameters) that the RF model identifies in order to make the improved mass estimates.

Some of the morphological parameters employed, particularly w and the power ratios, will be more difficult to measure accurately for *eROSITA*-observed clusters. Motivated by our finding that S, A, and c are the most important morphological parameters, we consider an additional RF model that is cross-validated and trained with a reduced feature set that includes only these three parameters and $L_{\mathrm{ex,z}}$. As displayed in Table 6.2, we find that this reduced model yields a 1σ scatter of $\delta = 0.070$ for the "idealized *Chandra*" series and $\delta = 0.071$ for the "realistic *eROSITA*" series. While this is still a 12–13% improvement over the mass-luminosity relation, this finding highlights the benefit of including the additional morphological parameters, even considering that each of them has an importance of $\lesssim$ 1%. The combined effect of the

additional morphological parameters, including e, w, M_{20}, and the power ratios, is ultimately responsible for roughly a third of the overall improvement offered by our approach.

6.6 Conclusion

In this work, we have presented a method for estimating cluster masses from mock X-ray observations of galaxy clusters. The mock observations are generated from 2,041 clusters with masses in the range of $10^{13.5} \leq M_{500c}/(h^{-1}\,\mathrm{M}_\odot) \leq 10^{14.8}$ and over a redshift range of $0.1 \leq z \leq 0.29$ from the `Box2` and `Box2b` *Magneticum* simulations. The mass predictor is based on a random forest regression model, trained with a feature set that includes only the core-excised luminosity and a set of morphological parameters, all of which can be computed directly from an X-ray image.

We demonstrate that this method can be used to estimate the masses of galaxy clusters with negligible bias and a scatter of $\delta = 0.066$ dex (16%). This model achieves a 20% reduction in the scatter relative to a more standard core-excised luminosity power law. Importantly, the same level of improvement is present both when using idealized, high-resolution, long-exposure (1 Ms) *Chandra* mock observations with no background and when using realistic, low-resolution, short-exposure (2 ks) *eROSITA* mock observations with added background noise. The majority of this improvement comes from three parameters: smoothness S, asymmetry A, and surface brightness concentration c. A more conservative model, which includes only the luminosity and these three parameters, estimates the cluster masses with a scatter of $\delta = 0.070$, demonstrating that a third of the overall improvement comes from the inclusion of the additional morphological parameters (e, w, M_{20}, and the power ratios). However, it is yet to be seen how additional sources of error present in real observations will affect the performance of this model; for example, the scatter in R_{500c} measurements will propagate to increased scatter in the morphological parameters.

While excising the cluster core reduces the scatter in mass estimates, this improvement comes at the cost of lowering the photon counts used in the analysis. However, even at the *eROSITA* detection threshold of ∼30 core-excised photon counts, the statistical uncertainty on $L_{\mathrm{ex,z}}$ will be $\lesssim$ 20%; hence, the mass estimate errors from a mass-luminosity relationship will be dominated by intrinsic scatter even in the low-photon limit. Since the dynamical state of the cluster encodes important information that affects mass errors, the inclusion of morphological parameters, which utilize the full photon distribution, in the mass model enable more accurate predictions with reduced intrinsic scatter relative to $M - L_{\mathrm{ex,z}}$. However, we expect that the statistical uncertainty in the mass estimates will still closely follow the corresponding statistical uncertainty of $L_{\mathrm{ex,z}}$. The relationship between the morphological parameters, the luminosity, and the mass is complicated and nonlinear; we have demonstrated that the nonlinear RF regression method

offers a substantial improvement over linear models.

Our model was trained to predict the spherical overdensity masses, M_{500c}, of galaxy clusters identified in *Magneticum* by the SUBFIND algorithm. As demonstrated in the Knebe et al. (2011) halo finder comparison project, this algorithm is able to estimate the M_{200c} masses of NFW host haloes to within $\lesssim 3\%$; more broadly, all modern halo finders compared in Knebe et al. (2011) are able to determine host halo M_{200c} to within $\lesssim 10\%$. Hence, we expect that uncertainty introduced due to the mass estimates of our simulated clusters is sub-dominant, but not insignificant, compared to the intrinsic scatter of the $M - L_{\mathrm{ex,z}}$ relationship. The recent *FABLE* simulations project (Henden et al., 2018), a set of hydrodynamical simulations with similar subgrid physics to *Magneticum*, reports that the $M - L_X$ of simulated clusters is in excellent agreement with the observed relation based on X-ray hydrostatic masses. Thus, if there is indeed an X-ray hydrostatic mass bias, this would indicate that simulated clusters may be too gas-rich, resulting in L_X values that are high relative to weak lensing-calibrated $M - L_X$. In addition, current models of AGN feedback result in simulated cluster cores that do not match the observed cool-core and non-cool-core cluster populations. Because of this, the morphological parameters of observed clusters that depend most sensitively on the cluster core (i.e., X-ray surface brightness concentration) may be biased relative to observations. We expect that these simulation sources of uncertainty will improve as the hydrostatic mass bias quandary approaches a resolution and as more sophisticated AGN feedback models are developed.

Random forest models are notoriously bad at extrapolation. We expect that our model's systematic underprediction of the masses of clusters in the high-mass tail of the halo mass function will improve when trained on a cluster sample that is uniform across the full mass range of interest. When large-volume, high-resolution hydrodynamical simulations become available for creating such a training sample, we expect that this can be used to train a model that predicts well across the entire mass range. This model, once trained on a sufficiently large simulated sample, could then be applied to a set of *Chandra*-observed clusters, such as the HIFLUGCS sample (Zhang et al., 2011), and the predictions could be tested against accurate mass estimates, such as those based on Y_X.

ML-based methods of estimating galaxy cluster masses from X-ray observations, including the method presented here as well as others in the literature (e.g., Ntampaka et al., 2019a), offer a promising step towards extracting the maximum information content present in imminent datasets such as *eROSITA*. Modern ML methods will enable the completion of an unprecedentedly accurate cosmic census and position the halo mass function to be used to place ever stronger cosmological constraints. Ultimately, the continued progress in cosmological hydrodynamical simulations, both in terms of physical realism and size, are rapidly facilitating the coming of age of these techniques, which will soon be ready for deployment on state-of-the-art cluster observation samples.

Chapter 7

Conclusion

THE growth and evolution of dark matter structure, on both large and small scales, is highly dependent upon the underlying particle nature of dark matter. Thus, observational efforts to infer the abundance of dark matter substructure (i.e., subhaloes) can be used to constrain the dark matter model parameter space. This is, however, only possible if we are able to make accurate predictions of the expected subhalo abundance given a particular dark matter model. The tools of the trade used to make such predictions — cosmological N-body simulations, full-physics cosmo-hydrodynamical simulations, and semi-analytical models — each come with their own set of limitations (e.g., computational, numerical, statistical, or physical realism). The work laid out in this dissertation represents a substantial leap towards quantifying the current numerical limitations of cosmological N-body simulation-based approaches to substructure prediction. To this end, we have built a highly accurate semi-analytical model of substructure that is not adversely impacted by the numerical limitations of such simulations. This model, `SatGen`, is fit to address a range of scientific questions, several of which we have already tackled. In what follows, we summarize the key conclusions from each of the studies that constitute this body of work.

A tidally evolved subhalo is determined by its initial state and the fraction of mass lost since infall.

- Using small suites of idealized N-body simulations, Hayashi et al. (2003) and Peñarrubia et al. (2008) developed simple models that capture the structural evolution of a NFW subhalo due to the tidal forces of its host. Common to their distinct models is the notion of a "tidal track", which is based on the empirical finding that the structure of a tidally evolved subhalo only depends on the fraction of its initial bound mass that has been lost since infall.

- In (Ogiya et al., 2019), we introduced the *DASH* simulation database, which includes 2,253 idealized, high-resolution N-body simulations of individual subhaloes orbiting within a static, analytical host potential. These simulations span the parameter space of possible subhalo orbits as well as the host- and subhalo concentrations. The wealth of data available in this simulation library has enabled us to calibrate a new model of the tidally evolved subhalo structure.
- We found that the residuals of the Hayashi et al. (2003) and Peñarrubia et al. (2008) evolved density profile models correlate strongly with the initial subhalo concentration, $c_{\rm s}$. This limitation was not identified in the previous studies because the authors only considered subhaloes of a single $c_{\rm s}$. Our model incorporates this additional $c_{\rm s}$-dependence.
- We presented our model of the evolved subhalo density profile (and structural parameters, $V_{\rm max}$ and $r_{\rm max}$) in Green & van den Bosch (2019, presented in chapter 2). This model is both more accurate and more general than its predecessors — it was calibrated over a wide range of subhalo parameter space and is sufficiently complex to perform well over this range at capturing the non-linear tidal evolution.

Cosmological N-body simulations still suffer from artificial disruption.

- In Jiang et al. (2021), we introduced a full dynamical semi-analytical model of the build-up and evolution of dark matter subhaloes and satellite galaxies, `SatGen`. In its present form, the model incorporates (i) analytical halo merger trees (Parkinson et al., 2008), (ii) a recipe for initializing subhalo orbits at infall (Li et al., 2020), (iii) an orbit integrator and a prescription for dynamical friction (Chandrasekhar, 1943), and (iv) a *DASH*-calibrated tidal evolution model (the first component of which is introduced in Green & van den Bosch, 2019).
- In Green et al. (2021a, presented in chapter 3), we augmented `SatGen` by incorporating the *DASH*-calibrated tidal evolution model. In addition to the density profile model of Green & van den Bosch (2019), we developed a prescription for subhalo mass-loss (via tidal stripping) that faithfully reproduces the mass trajectories of *DASH* subhaloes.
- By studying the properties *at accretion* of subhaloes in the *Bolshoi* simulation (Klypin et al., 2011) that are subsequently disrupted, we built a simple model (based on Jiang & van den Bosch, 2016) that captures the statistical properties of artificial disruption. This disruption mechanism can be applied to `SatGen` results during post-processing in order to assess the overall significance of such disruption.
- We used the augmented `SatGen` model to predict subhalo mass functions, number density profiles, radial bias profiles, and substructure mass fractions for a wide range of host halo masses. We considered the impact of artificial disruption (captured via our disruption mechanism) as well as the

simulation mass limit (i.e., "withering").

- When artificial disruption and withering are both taken into account, `SatGen` predicts subhalo demographics that agree exquisitely well with *Bolshoi*.
- Artificial disruption suppresses the overall substructure mass fraction (within $r_{\rm vir}$) by ∼10%. This effect becomes more significant towards the halo center, where roughly 50% of substructure is depleted within $0.1r_{\rm vir}$. However, the halo-to-halo variance in these quantities is significantly larger than the mean difference between the "disruption-on" and "disruption-off" results.
- Splashback haloes (i.e., subhaloes that have fallen into the host in the past but currently lie outside $r_{\rm vir}$) make up roughly half of the total subhalo population. These subhaloes *must* be taken into account in semi-analytical modeling (such as in `SatGen`) in order to generate accurate abundance estimates. Orbit-averaged models are inherently limited by their inability to take into account such splashback systems.
- The radial bias in the subhalo number density profile (relative to the smooth background profile of the host) is an artifact of the simulation mass resolution limit. In `SatGen`, when we allowed subhaloes to evolve down to arbitrarily small $m/m_{\rm acc}$, we found that this radial bias is eliminated. Thus, in addition to spurious subhalo disruption, it is clear that the limited mass resolution of cosmological simulations still has a significant impact (both qualitative and quantitative) on the subhalo demographics.

A stellar disc depletes subhaloes via enhanced tidal stripping, rather than disc shocking.

- Measuring the impact of a stellar disc on the subhalo population has been the topic of a wealth of studies (e.g., D'Onghia et al., 2010; Peñarrubia et al., 2010; Garrison-Kimmel et al., 2017; Errani et al., 2017). In the simulation study by D'Onghia et al. (2010), the authors found that an analytical disc potential embedded within a N-body host halo efficiently destroys subhaloes near the host center, which they attribute to disc shocking.
- Using `SatGen`, in Green et al. (2021b, presented in chapter 4), we assessed the influence of a stellar disc in a manner that (i) allowed us to quantify the halo-to-halo variance in a way that has not been possible in previous studies and (ii) is free of simulation-based artificial disruption. In particular, we generated 10,000 merger trees for Milky Way-mas systems and evolved the subhaloes under a variety of different composite halo–disc potentials.
- We found that the mean subhalo mass function (within $r_{\rm vir}$) is suppressed by $\lesssim$ 10% when 5% of the host mass is contained within a central disc potential. When only considering subhaloes within 50

kpc of the host center, this becomes a ∼30% suppression. However, this mean disc-driven suppression is dominated by the halo-to-halo variance in the individual subhalo mass functions.

- We followed individual subhaloes as they evolved under different composite host potentials. By comparing the $z = 0$ subhalo mass in the halo-only case to halo–disc configurations with different disc mass fractions, we demonstrated that the presence of a central disc drives enhanced mass loss, the significance of which grows with decreasing orbital pericentric radius.
- We demonstrated that the overall level of disc-driven subhalo depletion is minimally dependent on any property of the disc aside from its final mass, which is in agreement with the similar finding of Garrison-Kimmel et al. (2017).
- Finally, we argued that, rather than disruption via disc shocking, the primary effect of a disc potential is to simply increase the central density concentration of the host, which drives excess mass loss. When this effect is coupled with a simulation mass resolution limit, the result is that an increased number of subhaloes will fall below this mass limit, appearing to have been "disrupted" by the disc. In reality, these systems should still survive and have simply become insufficiently massive to remain viable in the simulation.

The three studies summarized above (Green & van den Bosch, 2019; Green et al., 2021a,b) constitute the primary research thrust of this dissertation. Using `SatGen`, we have demonstrated that the halo-to-halo variance in many quantities of astrophysical interest is quite large and, thus, must be properly taken into account when, e.g., placing limits on dark matter particle models. In a tangentially related study, we further explored this theme of significant assembly history-driven halo-to-halo variance in the context of the Sunyaev–Zel'dovich effect scaling relation of galaxy clusters (Green et al., 2020, presented in chapter 5). Indeed, variation in mass accretion histories also presents itself in a variety of other observables, such as the gas concentration, clumpiness, and spherical symmetry, which can be exploited to better infer halo masses of galaxy clusters (Green et al., 2019, presented in chapter 6).

Chapter 8

Future work

THE introduction of `SatGen` makes possible a variety of follow-up research programs. In what follows, we motivate and describe several such potential projects:

- In van den Bosch et al. (2016), the authors study the segregation of dark matter subhaloes from the *Bolshoi* simulation. The work identifies relationships between segregation indicators (e.g., host-centric radius, projected radius, and binding energy), which capture the relationship between a subhalo and its host, and segregation properties (e.g., m, $m_{\rm acc}$, $m/m_{\rm acc}$, $z_{\rm acc}$, $V_{\rm max}$, $V_{\rm acc}$, among others), which capture internal details about the subhalo and its history. As is clear from the study, there are rich segregation relationships between subhaloes and their hosts, at least as seen in cosmological simulations. For example, subhaloes are strongly segregated by accretion redshift, $z_{\rm acc}$; this is identified via a strong Spearman correlation between $z_{\rm acc}$ and binding energy. After having demonstrated in Green et al. (2021a) that `SatGen` accurately reproduces the evolved subhalo mass function and radial abundance profile of *Bolshoi* haloes when the statistical artificial disruption mechanism is enabled, a logical follow-up to this work is to perform a more comprehensive analysis of subhalo segregation as predicted by `SatGen`. In particular, it will be illuminating to study how well the aggregate of segregation relationships agree between `SatGen` and *Bolshoi* when the Jiang & van den Bosch (2016) artificial disruption mechanism is applied to the subhalo sample, as well as how significant of an impact disruption has on subhalo segregation. If any stark disagreements between disruption-enabled `SatGen` and *Bolshoi* are evident, this may provide further insight into how artificial disruption comes about in cosmological simulations and how it could be properly modeled (either physically or via machine learning, as in Nadler et al. 2018), mitigated and eventually corrected for.

- Recently, Nadler et al. (2021) carried out a study that used observations of Milky Way satellite galaxies

to place very strict bounds on the warm dark matter (WDM), self-interacting dark matter, and fuzzy dark matter models. In Jiang et al. (2021), we use the full `SatGen` framework, which assumes a model of the galaxy-halo connection to populate and evolve satellite galaxies within the DM subhaloes, to demonstrate that there is an extreme level of halo-to-halo variance in the satellite galaxy cumulative $v_{\rm max}$ (maximum circular velocity) distribution and radial abundance profile of Milky Way-mass systems. We suspect that a careful consideration of the predicted overall halo-to-halo variance and the scatter in the galaxy-halo connection for low-mass satellites is likely to weaken bounds placed on the particle nature of dark matter reported in the literature.

Using WDM as a case study, we propose the following test. First, we would run additional *DASH*-like simulations, this time varying the inner slope of the subhalo density profile in order to test how well the Green & van den Bosch (2019) ESHDP model and Green et al. (2021a) mass-loss model is able to reproduce subhalo evolution trajectories for non-NFW subhaloes. If necessary, we would pause to develop a more general ESHDP model that works for a range of initial subhalo density profiles and implement this into `SatGen`. We would then use `SatGen` to generate two, large-sample sets of subhalo and satellite galaxy catalogs for Milky Way-mass systems. The first would be based on analytical merger trees using the standard CDM power spectrum, whereas the second would use a WDM power spectrum (Viel et al., 2005) for a WDM particle mass somewhat below the lower limit of 6.5 keV set by Nadler et al. (2021). The cuspiness of the initial DM subhalo density profiles at infall would take into account the halo response to baryonic processes (Freundlich et al., 2020b); this flexible model has been calibrated to match the halo response seen in both strong-feedback simulations (e.g., FIRE; Hopkins et al., 2018) and weak-feedback simulations (e.g., NIHAO; Wang et al., 2015). Satellite galaxies can be populated and evolved using a model of the galaxy-halo connection currently built into SatGen; however, we could consider collaborating with the authors of `UniverseMachine` (Behroozi et al., 2019) in order to use their state-of-the-art framework for painting realistic satellite galaxies onto subhaloes. This procedure would enable us to predict the average satellite galaxy luminosity functions and halo-to-halo variance for Milky Way-mass systems in both CDM and WDM. Ultimately, we would expect to be able to demonstrate that when considering a realistic level of halo-to-halo variance, WDM models with a WDM particle mass considerably below current lower limits should still be able to reproduce the observed Milky Way satellite galaxy luminosity function.

- The presence of substructure can significantly boost the dark matter annihilation rate within an otherwise smooth host halo. Many previous numerical and analytical works have attempted to pin down the magnitude of this "boost factor", with most interest being directed towards local satellite galaxies

and massive galaxy clusters (for a recent review, see Ando et al., 2019). The boost factor is sensitive to (i) the internal density structure of evolved subhaloes, (ii) the minimum subhalo mass (set by the dark matter free-streaming scale), and (iii) the slope and normalization of the evolved subhalo mass function. Using `SatGen`, we could determine the mass- and redshift-dependence of the subhalo mass function. Additionally, we could use `SatGen` to establish how the distribution of subhalo bound mass fractions, $m/m_{\rm acc}$, depends on sub-to-host mass ratio, m/M, and host mass, M. The Green & van den Bosch (2019) evolved subhalo density profile is set by $m/m_{\rm acc}$, and thus we could also determine the distribution of evolved subhalo density profiles at fixed m/M and M. In aggregate, we could make simulation-independent predictions for how the boost factor depends on host halo mass, redshift, and minimum subhalo mass and compare to previous studies in the literature.

Appendix A

The *DASH* database

This chapter has been published as an article by Ogiya et al. (2019) in the Monthly Notices of the Royal Astronomical Society by Oxford University Press. Since I am a co-author of this work, I have elected to include it as an appendix because it provides a more thorough description of the *DASH* simulation methods than that of Section 2.2.

A.1 Background

In the cold dark matter (CDM) paradigm of cosmic structure formation, smaller perturbations collapse first to form virialized dark matter haloes, leading to a hierarchical assembly of haloes. When dark matter haloes assemble their mass by accreting smaller haloes, they build up a hierarchy of substructure, with subhaloes hosting sub-subhaloes, hosting sub-sub-subhaloes, etc. As these subhaloes orbit their hosts, they experience mass loss due to the combined effect of dynamical friction, tidal stripping and impulsive (tidal) heating (e.g., Mo et al., 2010). The resulting abundance and demographics of substructure depends on the microscopic properties of the dark matter particles, most importantly the free-streaming scale and the strength of dark matter self-interaction (see e.g., Knebe et al., 2008; Rocha et al., 2013; Bose et al., 2017). This is why many efforts are underway to quantify the amount of dark matter substructure using, among others, gravitational lensing (e.g. Vegetti et al., 2012; Shu et al., 2015; Hezaveh et al., 2016), gaps in stellar streams (e.g. Carlberg, 2012; Ngan & Carlberg, 2014; Erkal et al., 2016), and annihilation or decay signals of dark matter particles (e.g. Strigari et al., 2007; Pieri et al., 2008; Hayashi et al., 2016; Hiroshima et al., 2018). In addition, substructure is also directly related to the abundance and properties of satellite galaxies (e.g., Vale & Ostriker, 2006; Newton et al., 2018), and thus to the clustering amplitude of galaxies on small scales (see e.g., Benson et al., 2001; Berlind et al.,

2003; Kravtsov et al., 2004). Hence, it is important that we are able to make accurate predictions for the abundance of dark matter substructure for a given cosmological model. Given the highly non-linear nature of the processes involved, this is ideally done using N-body simulations.

Modern, state-of-the-art cosmological N-body simulations predict that roughly 5-10 percent of a halo's mass is bound up in substructure, with more massive host haloes having a larger subhalo mass fraction (e.g., Gao et al., 2004; Giocoli et al., 2010). In addition, subhaloes are found to be spatially anti-biased with respect to the dark matter, in that the radial number density profile, $n_{\rm sub}(r)$, is less centrally concentrated than the halo density profile (e.g., Diemand et al., 2004; Nagai & Kravtsov, 2005; Springel et al., 2008). It has also been concluded that the subhalo mass function has a universal form (Jiang & van den Bosch, 2016), albeit with a significant halo-to-halo variance at the massive end (Jiang & van den Bosch, 2017).

Although these trends are well understood (van den Bosch et al., 2005b; Zentner et al., 2005a; Jiang & van den Bosch, 2016), and the results from numerical simulations seem to be well converged (e.g., Springel et al., 2008; Onions et al., 2012), some issues remain. Foremost among these is the fact that subhaloes in numerical simulations typically experience complete disruption some time after accretion (e.g., Han et al., 2016; van den Bosch, 2017). Although it is often argued that this disruption is a physical consequence of either tidal stripping or tidal heating (Hayashi et al., 2003; Taylor & Babul, 2004; Klypin et al., 2015), others have argued that in a collisionless dark matter simulation, subhaloes should rarely ever completely disrupt. In particular, in van den Bosch et al. (2018, hereafter Paper I), we have demonstrated that neither tidal stripping nor tidal heating is expected to be able to completely unbind the central cusps of CDM substructure. The same conclusion was reached by Peñarrubia et al. (2010) using idealized, high-resolution numerical simulations. In van den Bosch & Ogiya (2018, hereafter Paper II), we used a large suite of similar, idealized simulations to demonstrate that the disruption of subhaloes in N-body simulations is predominantly numerical, and triggered by two independent aspects: discreteness noise and inadequate force softening.

An important finding of Paper II is that this artificial, numerical disruption may elude standard convergence tests, in that the method that is typically used to scale the force softening with the mass resolution is inadequate to overcome these problems. Artificial disruption can potentially have far-reaching consequences. After all, unless we can make accurate, and above all reliable, predictions regarding the abundance and structure of dark matter subhaloes, we will forfeit one of the main handles we have on learning about the nature of dark matter. In addition, artificial disruption is also a serious road-block for the small-scale cosmology program, which often relies on numerical simulations to predict the clustering strength of galaxies on small scales. A prime example of this is subhalo abundance matching, which assigns 'mock' galaxies to subhaloes identified in numerical simulations in order to predict galaxy-galaxy correlation functions (e.g., Vale & Ostriker, 2006; Conroy et al., 2006; Guo et al., 2010; Hearin et al., 2013). Although one may

overcome the implications of artificial disruption by including 'orphan' galaxies (i.e., mock galaxies without an associated subhalo in the simulation), this seriously diminishes the information content of small-scale clustering, unless it is well understood how many orphans to add and where.

Unfortunately, it is not clear how to improve N-body codes such that the issue of artificial disruption can be avoided. As a consequence, it is difficult to gauge its potential impact on our predictions for the abundance of substructure. However, some insight can be gained from the semi-analytical models constructed by Jiang & van den Bosch (2016). Using accurate halo merger trees, this model uses a simple, orbit-averaged prescription for mass stripping to predict the evolved subhalo mass function. The overall normalization for the efficiency of mass stripping is calibrated by matching the model predictions to those from a high-resolution cosmological N-body simulation. In addition, the model includes a treatment of subhalo disruption, which is also calibrated to accurately reproduce the disruption in the simulation. If we use this model, but turn off the disruption (rather than disrupting the subhaloes, we continue to strip their mass), the resulting subhalo mass function is roughly a factor of two higher than with disruption (Green & van den Bosch, 2019). Hence, if all disruption is indeed artificial, and if the mass stripping model used by Jiang & van den Bosch (2016) is roughly correct, numerical simulations may have been underpredicting the amount of surviving substructure by a factor of two. This would have far-reaching consequences for many areas of astrophysics. For instance, this factor of two is exactly what is needed to solve the 'galaxy clustering crisis' in subhalo abundance matching discussed in Campbell et al. (2018).

In order to make more reliable predictions, we need to develop more sophisticated semi-analytical models for the evolution of dark matter substructure. Numerous studies in the past have been devoted to this (e.g., Taylor & Babul, 2001; Peñarrubia & Benson, 2005; Zentner et al., 2005a; Diemand et al., 2007; Kampakoglou & Benson, 2007; Gan et al., 2010; Pullen et al., 2014), but they all have one shortcoming in common: they all use the outcome of cosmological N-body simulations in order to calibrate one or more 'fudge' parameters in their model. And in doing so, their semi-analytical models inherit the shortcomings of the simulations; put differently, by construction the models are only as accurate as the simulations used for their calibration.

In an attempt to bypass this shortcoming, this chapter presents a large database (called DASH, for Dynamical Aspects of SubHaloes), of more than 2,000 idealized, high-resolution simulations of the tidal evolution of individual subhaloes. The simulations cover most of the parameter space (mass ratios, orbital parameters and halo concentrations) relevant for modelling the tidal evolution of subhaloes as they orbit their hosts, and each simulation is evolved with sufficient numerical resolution that discreteness noise and force softening do not adversely affect their outcome. The primary goal of DASH is to enable a more accurate calibration and validation of analytical treatments of tidal stripping and heating, thereby allowing

for the construction of new and improved semi-analytical models for dark matter substructure that are not hampered by artificial disruption. In addition, we provide a non-parametric model of subhalo mass evolution, using random forest regression, along with the simulation data. This model describes the simulation results at the ~ 0.1 dex level and can be readily used in further modelling.

This chapter is organized as follows: Section A.2 describes the simulation setup and gives an overview of the DASH library. Section A.3 presents a few examples, highlighting the type of data that is available. In Section A.4, we demonstrate and validate the performance of a random forest regression model trained on the simulation data to predict bound mass fractions as a function of time since accretion for given orbital parameters and given concentrations of the sub- and host haloes. Finally, Section A.5 summarizes the chapter and discusses the future outlook.

A.2 Overview of the simulations

This section presents an overview of the DASH library of idealized, collisionless N-body simulations of halo minor mergers. After describing the initial conditions (Section A.2.1.1), the simulation code (Section A.2.1.3), and the analysis and products (Section A.2.1.4), we describe the method used to sample the parameter space (Section A.2.2) and the data format of the DASH library (Section A.2.3).

A.2.1 Simulation setup

A.2.1.1 Halo profiles and merger set-up

Each of the simulations run as part of the DASH library follows an individual N-body subhalo as it orbits the fixed, external potential of a host halo. Both the host halo and the *initial* (prior to the onset of tidal stripping) subhalo are assumed to be spherical, and to have a Navarro-Frenk-White (NFW; Navarro et al., 1997) density profile

$$\rho(r) = \rho_0 \left(\frac{r}{r_\mathrm{s}}\right)^{-1} \left(1 + \frac{r}{r_\mathrm{s}}\right)^{-2} , \tag{A.1}$$

where r_s and ρ_0 are the characteristic scale radius and density, respectively. We define the virial radius, r_vir, as the radius inside of which the average density is $\Delta_\mathrm{vir} = 200$ times the critical density given by $\rho_\mathrm{crit} = (3\,H_0^2/8\pi G)$, where H_0 and G are the Hubble constant and gravitational constant, respectively. The virial mass of the halo is given by

$$M_\mathrm{vir} = \frac{4\pi}{3}\,\Delta_\mathrm{vir}\,\rho_\mathrm{crit}\,r_\mathrm{vir}^3 . \tag{A.2}$$

We emphasize, though, that the DASH simulations also apply to other values of $\Delta_{\rm vir}$, as detailed in Section A.2.2 and Section A.2.1.2. The halo concentration is defined as $c \equiv r_{\rm vir}/r_{\rm s}$, and the virial velocity is defined as the circular velocity at the virial radius, $V_{\rm vir} = \sqrt{GM_{\rm vir}/r_{\rm vir}}$. The mass ratio between the host halo and the initial subhalo is specified by $\mathcal{M} \equiv M_{\rm vir,h}/M_{\rm vir,s}$, where, as throughout this chapter, subscripts 'h' and 's' indicate properties of the host- and subhaloes, respectively.

Initial conditions are generated under the assumption that the NFW subhalo is isolated and has an isotropic velocity distribution, such that its phase-space distribution function (DF) depends only on energy, i.e., $f = f(E)$. We use the method of Widrow (2000) to sample particles from the DF using the standard acceptance-rejection technique (Press et al., 1992; Kuijken & Dubinski, 1994). When computing the DF, we follow Kazantzidis et al. (2006) and assume that the initial NFW subhalo has an exponentially decaying density profile for $r > r_{\rm vir,s}$. We assume that the total mass in this exponential extension is $0.05M_{\rm vir,s}$, where $M_{\rm vir,s} = M(< r_{\rm vir,s})$. Requiring a smooth transition in the density profile at $r = r_{\rm vir,s}$ then determines the scale radius of the exponential decay. Note, though, that when we sample particles from the DF thus computed, we apply a hard truncation at $r = r_{\rm vir,s}$ (i.e., no particles are sampled beyond the subhalo's virial radius). Consequently, the system will deviate somewhat from perfect equilibrium near the truncation radius. However, since we embed the subhalo in an external tidal field, with a corresponding tidal radius that typically lies well inside of $r_{\rm vir,s}$, there is little virtue to having a subhalo whose outskirts are in perfect equilibrium. In fact, one might argue that it is more realistic to truncate the subhalo at the tidal radius corresponding to its initial position within the host halo. However, in addition to the tidal radius being ill-defined (see Paper I for discussion), we have demonstrated in Paper II that none of this matters; the simulation outcome is insensitive to whether we truncate the subhalo at the virial radius or at the initial tidal radius.

Throughout we adopt model units in which the gravitational constant, G, the initial virial radius of the subhalo, $r_{\rm vir,s}$, and the initial virial mass of the subhalo, $M_{\rm vir,s}$, are all unity. With this choice, the initial virial velocity of the subhalo, $V_{\rm vir,s}$ is unity, while the host halo has $V_{\rm vir,h} = \mathcal{M}^{1/3}$. Both have the same crossing time, $t_{\rm cross} \equiv r_{\rm vir}/V_{\rm vir} = 1$. In physical units, the crossing time is

$$t_{\rm cross} = 0.978h^{-1}\ {\rm Gyr}\left(\frac{\Delta_{\rm vir}}{200}\right)^{-1/2}. \tag{A.3}$$

where $h = H_0/(100\ {\rm km\,s^{-1}\,Mpc^{-1}})$. In what follows, whenever we quote time scales in physical units, we adopt $h = 0.678$ (Planck Collaboration et al., 2016a) and $\Delta_{\rm vir} = 200$, which implies that a time interval of $(\Delta t)_{\rm model} = 1$ corresponds to 1.44 Gyr. Note, though, that one can scale all the physical time scales quoted

in this chapter to other values of H_0 and $\Delta_{\rm vir}$ by simply multiplying the values quoted by the factor

$$\Gamma(h, \Delta_{\rm vir}) \equiv \left(\frac{h}{0.678}\right)^{-1} \left(\frac{\Delta_{\rm vir}}{200}\right)^{-1/2} . \tag{A.4}$$

Put differently, the mapping between time scales in DASH model units and physical units, is given by: $(\Delta t)_{\rm physical} = 1.44\ {\rm Gyr}\, \Gamma(h, \Delta_{\rm vir})\ (\Delta t)_{\rm model}$.

The DASH simulations span a wide range in orbital energy, E, and angular momentum, L. For convenience, we characterize the orbits using the following two dimensionless quantities:

- $x_{\rm c} \equiv r_{\rm c}(E)/r_{\rm vir,h}$, the radius of the circular orbit corresponding to the orbital energy, E, expressed in terms of the virial radius of the host halo.
- $\eta \equiv L/L_{\rm c}(E)$, the orbital circularity, defined as the ratio of the orbital angular momentum, L, and the angular momentum $L_{\rm c}(E)$ corresponding to a circular orbit of energy, E. Radial and circular orbits have $\eta = 0$ and 1, respectively.

We initially position the subhalo at the apocentre of its orbit, and follow its dynamical evolution for a period of $T_{\rm sim} = 36$ Gyr. The orbit's radial period is given by

$$T_{\rm r} = 2 \int_{r_{\rm p}}^{r_{\rm a}} \frac{{\rm d}r}{\sqrt{2[E - \Phi_{\rm h}(r)] - L^2/r^2}} , \tag{A.5}$$

(e.g., Binney & Tremaine, 2008), with $r_{\rm p}$ and $r_{\rm a}$ the pericentric and apocentric radii of the orbit, respectively, and $\Phi_{\rm h}(r)$ the gravitational potential due to the host halo. The latter is given by

$$\Phi_{\rm h}(r) = -V_{\rm vir,h}^2 \frac{\ln(1 + c_{\rm h} x)}{f(c_{\rm h})\, x} \tag{A.6}$$

where $x = r/r_{\rm vir,h}$ and

$$f(c) = \ln(1 + c) - c/(1 + c) . \tag{A.7}$$

Fig. A.1 plots $T_{\rm r}$ as function of $x_{\rm c}$ for several values of η and $c_{\rm h}$. Note that, to good approximation, $T_{\rm r} \sim 6.7\ {\rm Gyr}\, x_{\rm c}^{1.15}$, with only a very weak dependence on η or $c_{\rm h}$. Hence, for the range of orbits covered by DASH, which have $x_{\rm c} \in [0.5, 2.0]$ (see Section A.2.2 below), $T_{\rm sim} = 36$ Gyr corresponds to between 2.5 and 12 radial periods. The initial (Cartesian) vectors of the subhalo position and velocity with respect to the host halo are given by $\mathbf{X} = (r_{\rm a}, 0, 0)$ and $\mathbf{V} = (0, L/r_{\rm a}, 0)$, respectively, such that the subhalo orbit is confined to the $x - y$ plane. For each simulation we output a total of 301 snapshots, with the time interval between snapshots fixed at 0.12 Gyr.

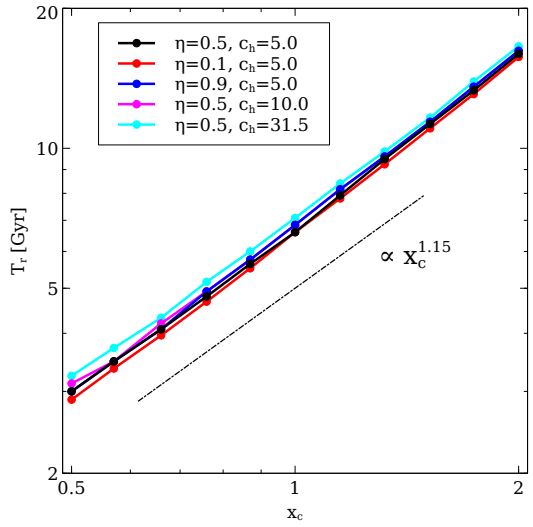

Figure A.1: The radial period, T_r, as a function of the parameter that controls the orbital energy of the subhalo, x_c. A virial overdensity of $\Delta_{vir} = 200$ and redshift of $z = 0$ are assumed. T_r scales as $T_r \propto x_c^{1.15}$, and it is almost independent of η and c_h.

A.2.1.2 Simulation Invariance

As discussed in Section A.2.1.1, the DASH simulations have been run in model units for which $G = M_{vir,s} = r_{vir,s} = 1$. In these units, the subhalo has a crossing time $t_{cross} \equiv r_{vir,s}/V_{vir,s}$ of unity. Converting to physical time units, simply requires multiplying the model time units by $1.44\,\Gamma$ Gyr with $\Gamma = \Gamma(h, \Delta_{vir})$ given by Equation (A.4). Hence, each DASH simulation is applicable to any combination of Δ_{vir} and $H_0 = 100\,h$ km s^{-1} Mpc^{-1}, each of which with its own scaling between time in model units and time in physical units.

Changing Δ_{vir} and/or h, while keeping the simulation parameters $(\mathcal{M}, x_c, \eta, c_h, c_s)$ fixed, corresponds to changing the actual physical densities. For example, increasing Δ_{vir} implies that the virial radius of the host halo decreases, such that a given value of x_c corresponds to a smaller physical radius, where the density of the host halo is larger. But, since the densities of the subhalo change similarly, and since the tidal evolution only depends on the ratio of densities (i.e., gravity is scale-free), the outcome of the simulation is entirely invariant to these changes in Δ_{vir} and/or h.

However, when changing Δ_{vir} and/or h one can also re-scale the DASH simulations in another way, one that keeps the physical densities, and hence the mapping between model time and physical time, invariant. This scaling, however, requires a mapping between the parameter set $\{\mathcal{M}, x_c, \eta, c_h, c_s\}$ for the DASH simulation ($\Delta_{vir} = 200$ and $h = 0.678$) and another parameter set $\{\mathcal{M}', x_c', \eta', c_h', c_s'\}$ corresponding to Δ_{vir}' and h'. This scaling keeps the characteristic density and scale-radius, i.e., ρ_0 and r_s in equation (A.1), invariant. Consequently, a different value of Δ_{vir}, which corresponds to a different virial radius, now implies a

different value for the mass and concentration parameter of the halo. And since the orbital radius remains invariant, also the dimensionless parameter x_c will change its value. Using that the characteristic density of an NFW halo with concentration parameter c is given by

$$\rho_0 = \frac{c^3}{f(c)} \Delta_{\rm vir} \rho_{\rm crit}, \tag{A.8}$$

it is straightforward to show that a DASH simulation for $\{\mathcal{M}, c_{\rm h}, c_{\rm s}, x_{\rm c}, \eta\}$ can be used to represent the evolution of a subhalo with parameters $\{\mathcal{M}', c'_{\rm h}, c'_{\rm s}, x'_{\rm c}, \eta'\}$ for any other combination of $\Delta'_{\rm vir}$ and $\rho'_{\rm crit}$ using the following mapping:

$$\begin{aligned}
\mathcal{M} &\rightarrow \mathcal{M}' = \mathcal{M}\,(\mathcal{Q}_{\rm h}^3/\mathcal{Q}_{\rm s}^3) \\
c_{\rm h} &\rightarrow c'_{\rm h} = c_{\rm h}\,\mathcal{Q}_{\rm h} \\
c_{\rm s} &\rightarrow c'_{\rm s} = c_{\rm s}\,\mathcal{Q}_{\rm s} \\
x_{\rm c} &\rightarrow x'_{\rm c} = x_{\rm c}\,\mathcal{Q}_{\rm h}^{-1} \\
\eta &\rightarrow \eta' = \eta \\
f_{\rm b} &\rightarrow f'_{\rm b} = f_{\rm b}\,(\Gamma/\mathcal{Q}_{\rm s}^3)\,.
\end{aligned} \tag{A.9}$$

Here $\Gamma = \Gamma(h', \Delta'_{\rm vir})$ is given by Equation (A.4), and $\mathcal{Q} = \mathcal{Q}(c, \Delta'_{\rm vir})$ is the root of

$$\mathcal{Q}^3 \frac{f(c)}{f(c\,\mathcal{Q})} \left(\frac{\Delta'_{\rm vir}}{200}\right) \left(\frac{h'}{0.678}\right)^2 = 1\,. \tag{A.10}$$

Note that, whereas the $(\Delta_{\rm vir}, h)$-dependent time-scaling is exact, this density-invariant mapping is only approximate. One of the reasons is that, because of dynamical friction, our simulations are only invariant to changes in $\mathcal{M}$ as long as the mass ratio $M_{\rm vir,h}/M_{\rm vir,s} \gtrsim 100$. Hence, one can only use this mapping as long as $(\mathcal{Q}_{\rm h}/\mathcal{Q}_{\rm s})^3 \gtrsim 0.1$. In addition, the initial subhalo is only initialized out to a truncation radius $r_{\rm trunc}$ (see Section A.2.1.1), which is equal to the virial radius, but only for $\Delta_{\rm vir} = 200$ and $h = 0.678$. For other values of $\Delta_{\rm vir}$ and h, the subhalo is effectively truncated at a radius $r_{\rm trunc} = r'_{\rm vir}/\mathcal{Q}_{\rm s}$. As we have demonstrated in Paper II, the simulation outcome depends only very weakly on where exactly the initial subhalo is truncated, as long as it is outside of the initial tidal radius. The weak dependence mainly originates from 'self-friction' (see Section A.3.1), which depends on the amount of mass that is stripped from the subhalo, which is larger if the initial truncation radius is larger. Hence, as long as the impact of self-friction is weak, which is almost always the case, we expect the density-invariant mapping of Equation (A.9) to be reasonably accurate. We have verified that this is indeed the case by running a few simulations for different values of $\Delta_{\rm vir}$, and comparing the resulting $f_{\rm b}(t)$ to predictions from the re-scaled DASH simulations based on the mapping of Equation (A.9). Except for the time period prior to the first pericentric passage, we find this mapping to be

accurate at the few percent level. The reason why the mapping fails prior to the first-pericentric passage is simply that the mapping predicts that the initial subhalo starts out with a bound mass fraction $f_{\rm b}'$ that is not equal to unity. However, after first pericentric passage the subhalo is basically stripped down to the same physical radius as in the fiducial case, and the mapping of Equation (A.9) is reliable.

A.2.1.3 Numerical techniques

All DASH simulations have been carried out using a tree code (Barnes & Hut, 1986) specifically developed for graphics processing unit (GPU) clusters (Ogiya et al., 2013). The code uses CPU cores to construct octree structures of the N-body particles, while GPU cards are used to compute gravitational accelerations through tree traversal. We employ 1,048,576 particles in each simulation. Forces between particles are softened using a Plummer softening length $\epsilon = 0.0003 r_{\rm vir,s}$ (see Paper II) and the opening angle of the tree algorithm is set to $\theta = 0.7$. Orbit integration uses a leapfrog scheme with a global, adaptive time step $\Delta t = 0.2\sqrt{\epsilon/a_{\rm max}}$ (Power et al., 2003). Here $a_{\rm max}$ is the maximum, absolute value of acceleration among all particles at that time[1]. As we have demonstrated in Paper II, these numerical parameters are adequate to properly resolve the tidal evolution of subhaloes. In order to verify this, we have run a subset of our simulations at ten times better mass resolution (using $N \sim 10^7$ particles), which yields results that are indistinguishable from our nominal mass resolution.

A.2.1.4 Data analysis and products

For each simulation output we compute the bound mass fraction of the subhalo, $f_{\rm bound}(t)$, using the iterative method described in detail in Appendix A of Paper I. Briefly, the centre-of-mass position, $\mathbf{r}_{\rm com}$, and velocity, $\mathbf{v}_{\rm com}$, are computed using the five percent most bound particles. These quantities are subsequently used to compute the binding energy of each particle. This is iterated until the changes in $\mathbf{r}_{\rm com}$ and $\mathbf{v}_{\rm com}$ are smaller than $10^{-4} r_{\rm vir,s}$ and $10^{-4} V_{\rm vir,s}$, respectively, which typically requires 3-10 iterations. The bound fraction, $f_{\rm bound}(t)$, is then defined as the fraction of the original particles that at time t have a negative binding energy.

All simulations in the DASH library initially have sufficient numerical resolution to properly resolve the dynamics of the dark matter subhalo. However, as highlighted in Paper II, simulation results can become unreliable once the subhalo has lost a significant fraction of its initial mass due to tidal evolution. Paper II presented the two criteria that a subhalo in a numerical simulation needs to satisfy in order for its dynamical evolution to be reliable. The first criterion, which is motivated by the work of Power et al. (2003), tests

1. Typically the time step is fairly constant throughout the entire simulation, except for a small (typically factor ~ 2) decrease in Δt during a high-speed pericentric passage of the host halo.

whether the softening length is sufficiently small to resolve the relevant accelerations. It translates into a requirement for the bound mass fraction given by

$$f_{\rm bound}(t) > 1.79\,\frac{c_{\rm s}^2}{f(c_{\rm s})}\,\left(\frac{\epsilon}{r_{\rm vir,s}}\right)\left(\frac{r_{\rm h}(t)}{r_{\rm vir,s}}\right) \tag{A.11}$$

with $r_{\rm h}(t)$ the half-mass radius of the (bound part of) the subhalo at time t. Using that all DASH simulations adopt $\varepsilon = 0.0003 r_{\rm vir,s}$, we thus have that the simulation results need to satisfy

$$f_{\rm bound}(t) > \frac{0.054}{f(c_{\rm s})}\,\left(\frac{c_{\rm s}}{10}\right)^2\,\left(\frac{r_{\rm h}(t)}{r_{\rm vir,s}}\right)\,, \tag{A.12}$$

for the results to be deemed reliable. The second criterion is related to discreteness noise, and puts a constraint on the number of bound particles in the subhalo. In particular, it states that the number of bound particles needs to exceed

$$N_{\rm crit} \equiv 80 N^{0.2}\,. \tag{A.13}$$

with N the number of particles in the *initial* subhalo. Once the number of bound particles falls below this critical value, the subhalo starts to experience a runaway instability, triggered by discreteness noise, which quickly leads to its demise (i.e., artificial disruption). Since all DASH simulations have $N = 1,048,576$, we have that $N_{\rm crit} = 1,280$, which implies that the DASH simulations are only reliable for

$$f_{\rm bound}(t) > 1.22 \times 10^{-3}\,. \tag{A.14}$$

As discussed in Section A.2.3 below, the DASH library contains, for each output, the bound mass fraction and the half-mass radius of the subhalo, which the user can use to test whether the output satisfies both criteria. More than 99.5% of all the simulation outputs available in the DASH library satisfy both criteria (A.12) and (A.14).

A.2.2 The DASH parameter space

The simulations described above are characterized by six parameters: the mass and concentration of the host halo, $M_{\rm vir,h}$ and $c_{\rm h}$, respectively, the mass and concentration of the subhalo, $M_{\rm vir,s}$ and $c_{\rm s}$, respectively, and the orbital parameters $x_{\rm c}$ and η. In order to limit the numerical cost of sampling this six-dimensional parameter space without sacrificing the volume sampled, we adopt a strategy that devotes more computational efforts to the regions of parameter space with higher probability. In doing so, we are guided by physical considerations and results from cosmological simulations.

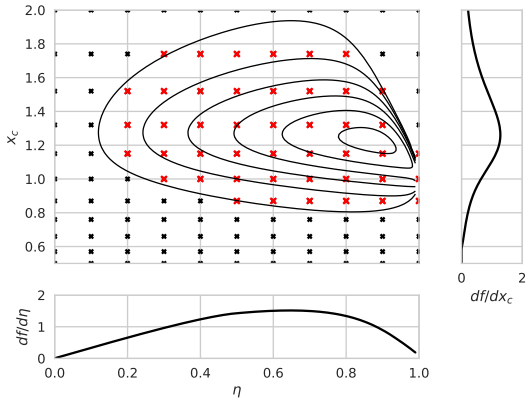

Figure A.2: Probability distribution of the orbital parameters, x_c and η, derived from the fitting function by Jiang et al. (2015) for a host halo mass of $M = 10^{12}\,M_\odot$ and minor merger mass ratios in the range $0.0001 \leq M_s/M_h \leq 0.005$. The host halo is assumed to follow a NFW density profile with a concentration $c_h = 5$. Crosses represent the parameter sets used in the DASH library. Red crosses cover the regions of parameter space with the largest probabilities in the distribution and most simulations are devoted to these regions. The right and bottom subset panels show the one dimensional marginalized probability distributions of x_c and η, respectively.

A.2.2.1 Initial mass ratio

Since dynamical timescales and the impact of tidal forces depend only on density, and since all haloes, by definition, have the same virial density, our simulation results should be independent of $M_{\rm vir,h}$ and $M_{\rm vir,s}$, significantly reducing the dimensionality of our parameter space. However, there is one important caveat. In reality, a subhalo orbiting a host halo experiences a dynamical drag force caused by the (perturbed) matter in the host halo ('dynamical friction'; e.g., Chandrasekhar, 1943) and by its own stripped material ('self-friction'; e.g., Fujii et al., 2006; Fellhauer & Lin, 2007). The impact of these drag forces, which result in orbital decay, increase strongly with decreasing mass ratio $\mathcal{M} = M_{\rm vir,h}/M_{\rm vir,s}$. When $\mathcal{M} \gtrsim 100$, though, their impact can safely be ignored (e.g., Mo et al., 2010). Throughout, we therefore restrict ourselves to simulations with $\mathcal{M} = 1000$. As we specifically demonstrate in Section A.3.1, these simulations accurately capture the tidal evolution of subhaloes for any $\mathcal{M} \gtrsim 100$, independent of the absolute value of $M_{\rm vir,h}$.

A.2.2.2 Sampling orbital parameters

Using a state-of-the-art cosmological simulation, Jiang et al. (2015) studied the orbital parameters of dark matter subhaloes at their moment of accretion into their host halo (see also e.g., Tormen, 1997; Zentner et al., 2005a; Khochfar & Burkert, 2006; Wetzel, 2011; van den Bosch, 2017). In particular, Jiang et al. (2015) measured the radial and tangential components, V_r and V_θ, of the relative velocity vector between the host- and subhaloes at infall. We convert their bivariate distribution of $V_r/V_{\rm vir,h}$ and $V_\theta/V_{\rm vir,h}$ to the corresponding bivariate distribution of x_c and η, assuming an NFW density profile for the host halo with a

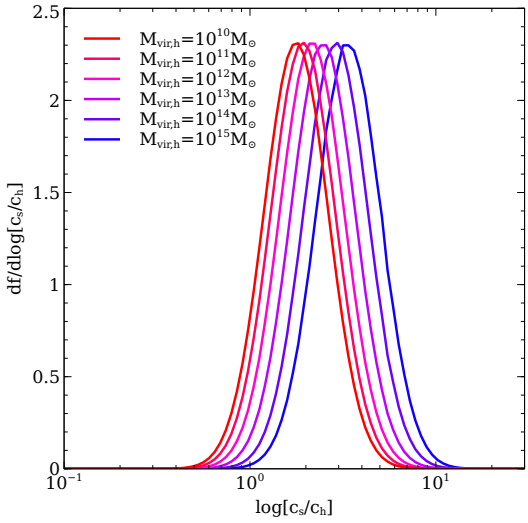

Figure A.3: Probability distribution of the halo concentration ratio, $c_\mathrm{s}/c_\mathrm{h}$, based on the halo merger rate by Fakhouri et al. (2010) and $c(M_\mathrm{vir}, z)$ relation by Ludlow et al. (2016), as well as the log-normal scatter of $\sigma_{\log c} = 0.12$ in the $c(M_\mathrm{vir}, z)$ relation. A redshift of $z = 0$ is assumed. Lines represent the distribution for host haloes with $M_\mathrm{vir,h} = 10^{10}$, 10^{11}, 10^{12}, 10^{13}, 10^{14} and 10^{15} $M_\odot$, respectively.

concentration parameter $c_\mathrm{h} = 5$. The resulting PDF $P(x_\mathrm{c}, \eta)$ is shown in Fig. A.2, where the small panels show the corresponding, marginalized distributions of x_c (side panel) and η (bottom panel).

In order to ensure that our sampling of x_c and η covers the entire range relevant for modelling the assembly and evolution of dark matter substructure, we proceed as follows. We sample η linearly over the entire range from $\eta = 0$ (purely radial orbit) to 1.0 (circular orbit) in 10 steps of $\Delta\eta = 0.1$, and x_c logarithmically over the range from $x_\mathrm{c} = 0.5$ to 2.0, in 10 steps of $\Delta \log x_\mathrm{c} \simeq 0.06$. Hence, there are a total of 121 combinations of (x_c, η). Together with 121 combinations of $(c_\mathrm{h}, c_\mathrm{s})$ (see below), this would imply that we need to run more than 14,500 simulations. In addition to being prohibitively expensive, this is also unwarranted, as we will demonstrate in Section A.4. Instead, we significantly reduce the number of simulations by focusing most of our efforts on the more likely parts of parameter space. The red crosses in Fig. A.2 indicate the 45 most likely combination of x_c and η. For each of these we run the 45 most likely combinations of $(c_\mathrm{h}, c_\mathrm{s})$, using the sampling strategy detailed below. The black crosses, on the other hand, indicate the combinations of (x_c, η) for which we only run one realization of halo concentrations, namely the one for $(c_\mathrm{h}, c_\mathrm{s}) = (5.0, 10.0)$.

A.2.2.3 Sampling halo concentrations

In deciding on how to sample c_h and c_s, we are again guided by cosmological simulations. These reveal that haloes of fixed mass have a log-normal distribution of halo concentrations, with a scatter of ~ 0.12 dex, and with a median that decreases with halo mass roughly as $c \propto M^{-0.1}$ (e.g., Bullock et al., 2001; Prada et al.,

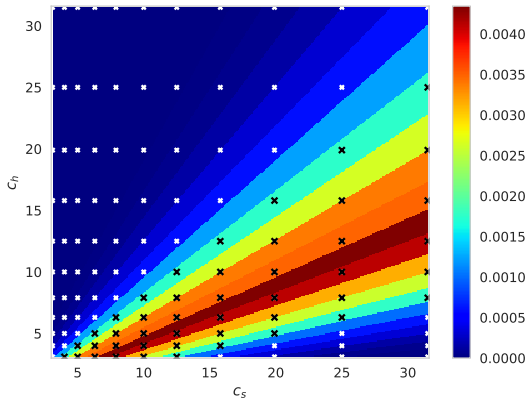

Figure A.4: Probability distribution of the halo concentration parameters, $c_{\rm h}$ and $c_{\rm s}$. The line for $M_{\rm vir,h} = 10^{12}\,M_\odot$ in Fig. A.3 is translated into this two-dimensional space, assuming that the distribution in Fig. A.3 is independent of $M_{\rm vir,h}$. Crosses represent the parameter sets studied in the DASH library. Black crosses cover the regions of largest probability in the distribution and most simulations are devoted to these concentration pairs.

2012; Dutton & Macciò, 2014; Diemer & Kravtsov, 2015; Ludlow et al., 2016). For DASH we sample both $c_{\rm s}$ and $c_{\rm h}$ using $\log c = 0.5, 0.6, 0.7, ..., 1.5$. This covers the range $3.1 \lesssim c_{\rm vir} \lesssim 31.5$, which is adequate for the vast majority of all haloes in our mass range of interest (roughly $10^7 < M_{\rm vir}/(\,h^{-1}{\rm M}_\odot) < 10^{15}$). This sampling, though, yields a total of 121 unique pairs of $(c_{\rm h}, c_{\rm s})$. Similar to our sampling of the orbital parameters, we focus most of our efforts on the 'most likely' combinations of $c_{\rm h}$ and $c_{\rm s}$. First, using the fitting function of Fakhouri et al. (2010), we compute the halo merger rate as a function of the mass ratio of the merging haloes. We only consider minor mergers with a mass ratio $\mathcal{M} > 100$, representative of our simulations. Next, we draw concentrations for each of the merging haloes using the concentration-mass relation of Ludlow et al. (2016), and account for log-normal scatter with $\sigma_{\log c} = 0.12$. Using this Monte-Carlo procedure, we compute the PDF for $\log[c_{\rm s}/c_{\rm h}]$. Results for different values of $M_{\rm vir,h}$ are shown in Fig. A.3, and are in good agreement with the results obtained from cosmological simulations (e.g., Paper I). Note that more massive host haloes tend to have somewhat higher concentration ratios on average. However, the strength of this mass dependence is weak compared to the width of the individual PDFs, and we therefore ignore it in what follows. Fig. A.4 shows the joint distribution of $c_{\rm h}$ and $c_{\rm s}$, derived using the PDF for $\log[c_{\rm s}/c_{\rm h}]$ for a host halo of mass $M_{\rm vir,h} = 10^{12}\,h^{-1}{\rm M}_\odot$ and adopting a uniform distribution[2] in $c_{\rm h}$. Black crosses indicate combinations of $(c_{\rm h}, c_{\rm s})$ for which we have run simulations for all 45 most likely orbital parameter combinations $(x_{\rm c}, \eta)$ (the red crosses in Fig. A.2). White crosses, on the other hand, indicate combinations of $(c_{\rm h}, c_{\rm s})$ for which we have run only one orbital configuration, namely with $(x_{\rm c}, \eta) = (1.0, 0.5)$.

2. Although halo concentrations follow a log-normal distribution at fixed mass, we adopt (for simplicity) a uniform distribution such that the simulations are relevant for a wide range in halo masses.

filename	description
`subhalo_evo`	position, velocity, bound mass fraction, and half-mass radius of subhalo as function of t
`radprof_rho`	subhalo density, $\rho(r,t)$, as function of r and t, in units of $200\rho_{\rm crit}(0)$
`radprof_m`	enclosed mass of subhalo, $M(<r,t)$, as function of r and t, in units of $M_{\rm vir,s}$
`radprof_sigmar`	radial velocity dispersion, $\sigma_{\rm r}(r,t)$, as function of r and t, in units of $V_{\rm vir,s}$
`radprof_sigmat`	tangential velocity dispersion, $\sigma_{\rm t}(r,t)$, as function of r and t, in units of $V_{\rm vir,s}$

Table A.1: Summary of the DASH data files available for each simulation.

A.2.3 Data structure

Using the sampling strategy outlined above, there are a total of 2,177 idealized N-body simulations that we have run. It is hierarchically structured, with different layers of directories corresponding to the various adopted orbital or concentration parameters. Each directory presents the results of the simulation in the form of five simple text files (see Table A.1), and each file contains a header listing the parameter set used in the simulation, as well as a brief description of the data file. Rather than accessing the results for individual simulations, the user can also download a single compressed archive file ($\sim 0.6\,$GB) containing the results for all 2,177 simulations as well as a Python routine that uses a non-parametric model based on random forest regression (see Section A.4), to predict the bound mass fraction of a subhalo as a function of time, for any configuration captured by our parameter space. In what follows we briefly describe the information that is provided for each simulation.

The file `subhalo_evo` lists the temporal evolution of several subhalo properties. The first column lists the ID of each simulation snapshot, which can be used to compute the corresponding physical time as

$$t = 0.12\ \mathrm{Gyr}\,\Gamma(h,\Delta_{\rm vir}) \times \mathrm{ID}\,. \tag{A.15}$$

Columns 2-4 and 5-7 list the Cartesian components of the position and velocity vectors of the centre-of-mass of the subhalo, respectively. The former and latter are normalized by the virial radius and velocity of the host halo, $r_{\rm vir,h}$ and $V_{\rm vir,h}$, respectively. Finally, columns 8 and 9 list the subhalo bound mass fraction, $f_{\rm bound}(t)$, and the subhalo's half-mass radius, $r_{\rm h}(t)$. The latter is normalized by the virial radius of the initial subhalo, $r_{\rm vir,s}$. Note that these outputs describe the tidal evolution of a (spherical NFW) subhalo of any mass, $M_{\rm vir,s}$, in a (spherical NFW) host halo of any mass, $M_{\rm vir,h}$, as long as $\mathcal{M} = M_{\rm vir,h}/M_{\rm vir,s} \gtrsim 100$. For smaller mass ratios dynamical friction, which is not accounted for, is important.

The other data available for each simulation in the DASH library are four text files that list the time evolution of four different radial profiles: the normalized density, $\rho(r,t)/[200\,\rho_{\rm crit}(0)]$, the enclosed, normalized mass profile, $M(<r,t)/M_{\rm vir,s}$, and the normalized radial and tangential velocity dispersion profiles, $\sigma_{\rm r}(r,t)/V_{\rm vir,s}$ and $\sigma_{\rm t}(r,t)/V_{\rm vir,s}$, respectively. The positions and velocities of all particles are defined

with respect to the centre-of-mass position and velocity of the subhalo, and only bound particles are taken into account. The first row of each of the `radprof` files lists the radial bins, which span the range $-2.95 \leq \log(r/r_{\rm vir,s}) \leq 0.95$ in equally spaced bins of width $\Delta \log(r/r_{\rm vir,s}) = 0.1$. The subsequent rows list the radial profiles for each of the 301 simulation snapshots, with row j corresponding to snapshot ID=$j-2$. The corresponding physical time is given by equation (A.15).

A.3 Some examples from the simulations

While classical dynamical friction is absent in the simulations, another source of friction caused by the mass stripped from the subhalo may alter the mass and orbital evolution of the subhalo in the tidal interactions. Section A.3.1 studies the validity condition that must be satisfied in order to neglect this friction force. The subsequent subsections present some examples of the DASH simulations, namely the mass evolution (Section A.3.2) and radial profiles (Section A.3.3) of the tidally stripped subhaloes.

A.3.1 Dependence on the initial mass ratio

As mentioned in Section A.2.1, the host halo is modelled as an analytical NFW potential and the initial mass ratio between the host- and subhaloes, $\mathcal{M} \equiv M_{\rm vir,h}/M_{\rm vir,s}$, is fixed at 10^3 for all simulations. In this subsection, we discuss the validity condition for these assumptions.

In actual mergers, friction forces may alter the orbit of subhaloes. The classical dynamical friction (Chandrasekhar, 1943) is caused by wakes formed in the density field of the larger host system due to the gravitational force of smaller objects, such as supermassive black holes, galactic bars, and dark matter subhaloes (e.g., Weinberg & Katz, 2007; Antonini & Merritt, 2012; Ogiya & Burkert, 2016). This kind of friction force is absent in our simulations because we employ an analytical host potential and the response in the density field of the host system due to the gravitational force of the subhalo is therefore not taken into account. The timescale of orbital decay due to this dynamical friction is roughly equal to $\mathcal{M}\tau_{\rm ff}$ (e.g., Mo et al., 2010), where $\tau_{\rm ff}$ is the free-fall time of the host halo, which is of the order of 2-3 Gyr at $z = 0$. Since all our simulations adopt $\mathcal{M} = 10^3$, is it clear that we can safely neglect the effects of dynamical friction, and thus use an analytical potential to model the host halo.

However, in addition to this 'classical' dynamical friction, which is not accounted for in our simulations, there is an additional friction force that may contribute, and which *is* included in our simulations. This force is due to the mass stripped from the subhalo itself, and we therefore refer to this force as 'self-friction'. Fig. A.5 shows an example of the tidal evolution of a subhalo in the DASH simulation with $x_{\rm c} = 1.0$, $\eta = 0.5$, $c_{\rm h} = 5$, $c_{\rm s} = 10$ and $\mathcal{M} = 10^3$. Time progresses from the top left to the bottom right, as indicated, while

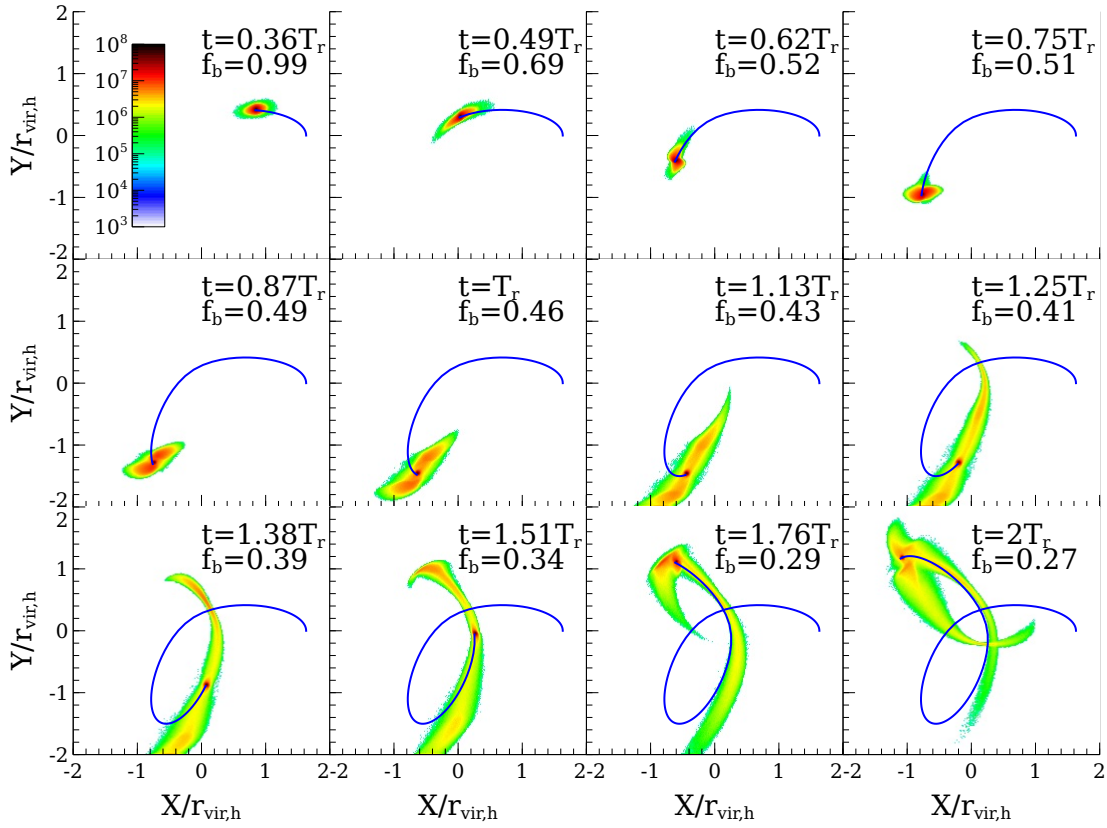

Figure A.5: Distribution of particles in the simulation run with $x_c = 1.00$, $\eta = 0.5$, $c_h = 5$, $c_s = 10$ and $\mathcal{M} = 10^3$. The blue line represents the subhalo orbit with respect to the centre of the host halo, which corresponds to the origin in this plot. From the leftmost panel in the top row to the rightmost panel in the bottom row, the evolution is illustrated. The time and bound mass fraction, f_b, of each snapshot are denoted in each panel. The spatial coordinates are scaled by the virial radius of the host halo, $r_{vir,h}$, and the colour bar (column density) is given in arbitrary units.

colors indicate column density (in arbitrary units). As is evident, the tidal field of the host halo strips mass from the subhalo, giving rise to a leading and a trailing tidal arm that roughly trace out the subhalo's orbit. The gravitational force from this stripped material on the bound remnant, can result in a net deceleration, thus giving rise to a self-friction (Fujii et al. 2006; Fellhauer & Lin 2007; Paper II).

In order to estimate the impact of self-friction, we have performed a number of simulations in which we vary the mass ratio, $\mathcal{M}$, while keeping all other parameters fixed to $x_c = 1.0$, $\eta = 0.5$, $c_h = 5$ and $c_s = 10$. Each subhalo is simulated with 1,048,576 particles and uses our fiducial softening and time-stepping. The results are shown in Fig. A.6, which plots the orbital radius (upper panel) and bound mass fraction (lower panel) as functions of time. The black dotted line corresponds to the results from a $\mathcal{M} = 10^3$ simulation with an order of magnitude more particles ($N = 10,485,760$), and is shown to demonstrate that these results are well converged; the black dotted line lies exactly on top of the results from our fiducial resolution simulation, shown as a purple solid line. In the absence of any friction, the results from all these simulations should all be identical. This follows from the scale-free nature of gravity, the fact that (sub)haloes with the same concentration parameter but different mass, have identical density profiles as function of the normalized radius r/r_{vir}, and the fact that the tidal radius scales with the ratio of the densities of host and subhalo[3].

While the simulations with $\mathcal{M} \gtrsim 100$ are indeed indistinguishable, those with smaller mass ratios clearly deviate, both in terms of their orbit and in terms of the bound mass fraction. Notably, the apocentric distance reached after the first pericentric passage becomes smaller for smaller $\mathcal{M}$, indicating that the subhalo is experiencing self-friction due to its own stripped material. Over time this self-friction causes the orbit to shrink, which reduces the pericentre of the orbit, thereby exposing the subhalo to a stronger tidal tidal, which in turn results in a reduced bound mass fraction. As is evident from Fig. A.6, the impact of self-friction is more pronounced for smaller values of $\mathcal{M}$. This is easy to understand; as $\mathcal{M}$ decreases, the force from the stripped material on the bound remnant *relative* to that from the host halo increases, causing a more pronounced deceleration along its orbit. We present a more detailed study of self-friction in Miller et al. (2020). Relevant for this study, though, is that self-friction is negligible for $\mathcal{M} \gtrsim 100$, and that the evolution of the bound mass fraction does not depend on $\mathcal{M}$ as long as this is the case. Hence, the DASH simulations presented here, which all have $\mathcal{M} = 10^3$, are valid for any mass ratio $\mathcal{M} \gtrsim 100$.

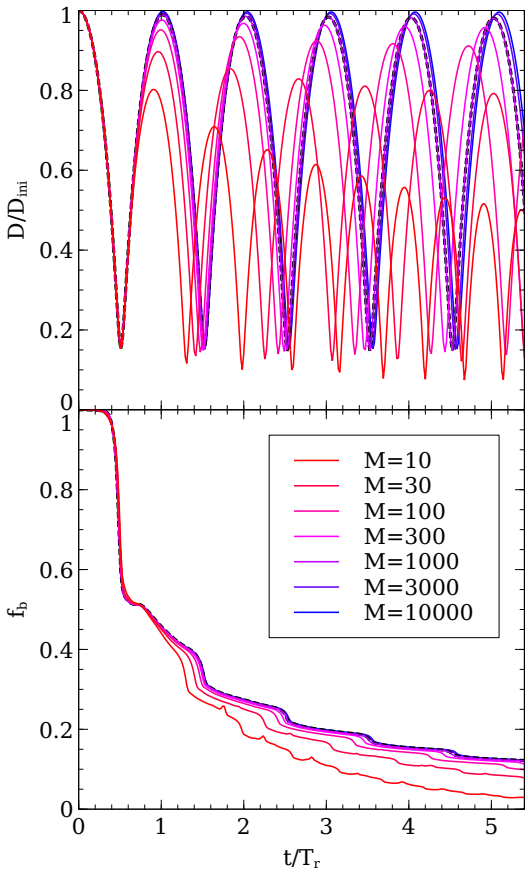

Figure A.6: (*Upper panel*) Distance between the centres of the host potential and subhalo, D, normalized by the initial value, $D_{\rm ini}$. (*Lower panel*) Bound mass fraction, $f_{\rm b}(t) \equiv M_{\rm b}(t)/M_{\rm vir,s}$, where $M_{\rm b}(t)$ and $M_{\rm vir,s}$ are the bound mass of the subhalo at each given time and the initial subhalo mass, respectively. The initial mass ratio between the host- and subhalo, $\mathcal{M} \equiv M_{\rm vir,h}/M_{\rm vir,s}$, is denoted in the legend. In the simulations, the same parameters for the host potential ($M_{\rm vir,h}$ is fixed and $c_{\rm h} = 5$) and subhalo orbit ($x_{\rm c} = 1.00$ and $\eta = 0.5$) are adopted. While the concentration of the subhalo is also fixed ($c_{\rm s} = 10$), the subhalo mass is varied and its spatial scale is altered accordingly (i.e., $r_{\rm vir,s} \propto M_{\rm vir,s}^{1/3}$). In the simulations represented by solid lines, the standard resolution level ($N \approx 10^6$) is adopted. As an additional test, the number of particles is increased by a factor of ten in the simulation represented by the black dotted line ($\mathcal{M} = 10^3$), with other parameters held fixed.

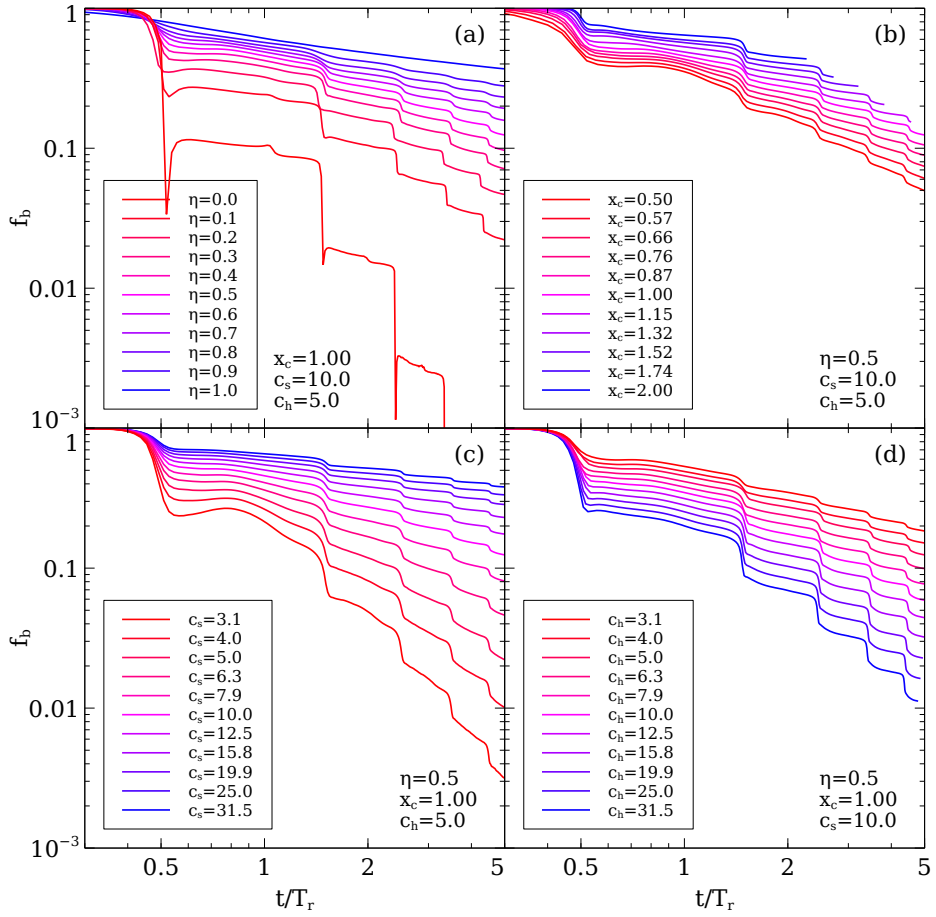

Figure A.7: Evolution of the subhalo bound mass fraction, f_b. The timescale is normalized by the radial period, T_r, of each orbit. In panels (a), (b), (c), and (d), each of the parameters η, x_c, c_s, and c_h are varied, respectively, holding the others fixed as denoted.

A.3.2 Bound mass fraction

Fig. A.7 demonstrates the evolution of the bound mass fraction of the subhalo, $f_{\rm b}$, obtained in some example simulations. Fixing the structural parameters of the host- and subhaloes (upper panels), the subhalo mass is more significantly decreased on more tightly bound (smaller $x_{\rm c}$) or radial (smaller η) orbits, since the orbits have smaller pericentres and the subhalo feels a stronger tidal force from the host potential. The lower panels illustrate that when the subhaloes are on the same orbit, a larger $c_{\rm h}$ (smaller $c_{\rm s}$) leads to more significant mass loss because of the stronger tidal force of the host halo (more loosely bound subhalo structure). These results, which only represent a tiny fraction of the entire DASH library, are consistent with intuition.

Less intuitive is the behavior in the bound mass fraction on a highly radial orbit (i.e., small η) close to a pericentric passage. In some cases, the bound mass fraction fluctuates wildly, dropping steeply, only to increase again immediately thereafter (see for instance the purely radial, $\eta = 0$ orbit in the upper-left panel). This arises i) because the method to compute the bound mass and orbit of the subhalo allows re-binding of particles (see Appendix A of Paper I for details), and ii) because the subhalo is impulsively heated by tidal shocking, especially on radial orbits (e.g., Spitzer, 1987; Gnedin et al., 1999b), which leads to the strong reduction of the bound mass. A fraction of the temporarily evaporated subhalo mass is re-bound during the subsequent re-virialization process of the subhalo.

Another interesting feature apparent in Fig. A.7 is that, except for the step-like behavior near pericentre, $f_{\rm b}$ roughly behaves like a power-law function for $t/T_{\rm r} > 1$. The slope of this power-law function clearly depends on the orbital and concentration parameters, and the DASH library, which contains simulation data for a large parameter space, can be used to calibrate these scaling relations, which in turn can be used to model the tidal evolution of substructure in (semi-)analytical models of structure formation. An alternative method, which we explore in Section A.4 below, is to use machine learning algorithms, such as random forest regression, to process and distill the huge amount of data available in the DASH library. As we demonstrate, this allows for reasonably accurate predictions for the bound mass fraction of subhaloes as a function of their time since accretion.

A.3.3 Radial profiles

The DASH library also includes density and velocity dispersion profiles of tidally stripped subhaloes. An example is shown in Fig. A.8, which plots the profiles for a subhalo for three snapshots of the simulation with $c_{\rm h} = 5.0$, $c_{\rm s} = 10.0$, $x_{\rm c} = 1.00$ and $\eta = 0.5$. Each of these snapshots corresponds to an epoch at

3. A comprehensive review is found in Paper I (see also e.g., Binney & Tremaine 2008)

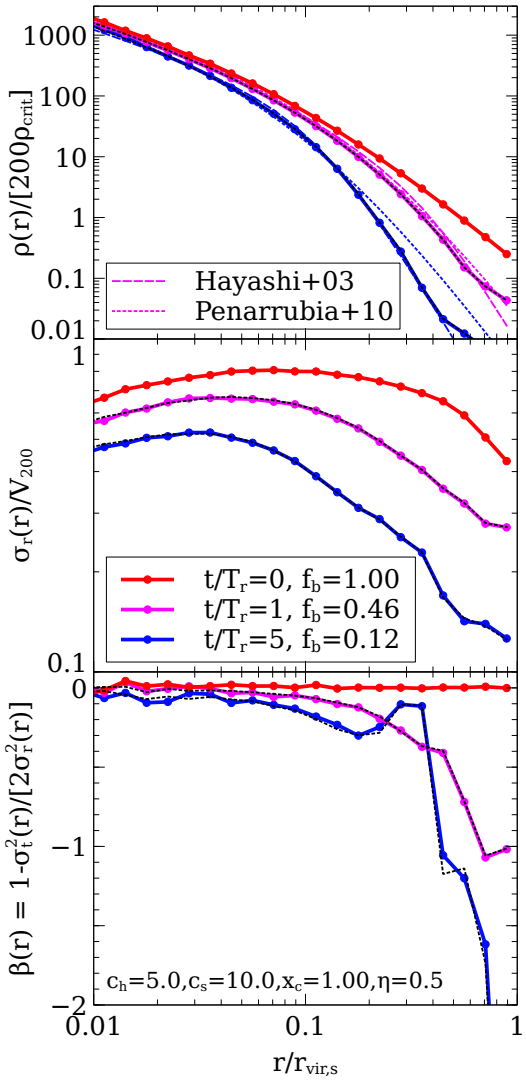

Figure A.8: Radial profiles of the dark matter subhalo in the DASH simulation with $c_{\rm h} = 5.0$, $c_{\rm s} = 10.0$, $x_{\rm c} = 1.0$, and $\eta = 0.5$. In each panel, coloured solid lines are the results from the simulation with the fiducial resolution and the black dashed lines demonstrate those from the higher resolution run (in which the number of particles is increased by a factor of 10, holding the other parameters fixed). (*Top*) Density profile. In addition to the simulation results, predictions by the models of Hayashi et al. (2003, coloured dashed) and Peñarrubia et al. (2010, coloured dotted) are shown. (*Middle*) Profile of the radial velocity dispersion, $\sigma_{\rm r}(r)$. (*Bottom*) Profile of the velocity anisotropy parameter, $\beta(r) \equiv 1 - \sigma_{\rm t}^2(r)/[2\sigma_{\rm r}^2(r)]$ (e.g., Binney & Tremaine, 2008), where $\sigma_{\rm t}(r)$ is the profile of the tangential velocity dispersion of the subhalo. The corresponding time and subhalo bound mass fraction are denoted in the middle panel.

which the subhalo is near its apocentre, at which point the subhalos is in a fairly relaxed state (e.g., Aguilar & White, 1986; Peñarrubia et al., 2009), and the corresponding bound mass fractions are indicated in the middle panel. The top panel of Fig. A.8 shows the subhalo's density profile, normalized by $200\rho_{\rm crit}$. Note how tidal stripping mainly removes mass from the outskirts, while leaving the central densities almost unaffected. For comparison, the dashed and dotted curves are the model predictions of Hayashi et al. (2003) and Peñarrubia et al. (2010), respectively. Both models suggest that the density profile of a stripped subhalo only depends on the initial density profile (prior to stripping) and the present bound mass fraction. Whereas the model by Hayashi et al. (2003) fits the profiles extremely well, the model by Peñarrubia et al. (2010) predicts a shallower outer density profile at later times. However, as we demonstrate in Green & van den Bosch (2019), neither the Hayashi et al. (2003) nor the Peñarrubia et al. (2010) model can adequately describe the evolution of the subhalo density profile under all conditions encountered in the DASH library, and we therefore develop a new and improved model based on the entire set of over 2,000 DASH simulations. The middle and bottom panels of Fig. A.8 show the radial velocity dispersion profiles and the corresponding profiles of the velocity anisotropy parameter

$$\beta(r) \equiv 1 - \frac{\sigma_{\rm t}^2(r)}{2\sigma_{\rm r}^2(r)}, \tag{A.16}$$

(Binney & Tremaine, 2008). Note that, by construction, the initial subhalo is isotropic ($\beta = 0$) throughout. At later times, as its bound mass fraction decreases, the subhalo becomes more and more tangentially anisotropic ($\beta < 0$) in its outskirts, while the radial velocity dispersion profile decreases on all scales. Hence, the bound remnant becomes colder and colder as more and more mass is stripped, and since particles on more radial orbits reach larger apocentric distances, they are more likely stripped, thereby causing the remnant to become tangentially anisotropic. Finally, the black, dashed lines show the results from a simulation with 10 times higher mass resolution. The fact that the resulting profiles are indistinguishable from those of the nominal resolution simulation indicates that the DASH simulations are well converged.

A.4 An application of machine learning to the DASH library

The DASH library is a homogeneously structured, large dataset that can easily be explored with machine learning (ML). As a first example, we apply a commonly-used ML regression method, random forests (RF; Breiman, 2001) as implemented in `scikit-learn` (Pedregosa et al., 2011)[4], to predict $f_{\rm b}$ as a function of five features: $t/T_{\rm r}$, $x_{\rm c}$, η, $c_{\rm h}$ and $c_{\rm s}$. Because $c_{\rm h}$, $c_{\rm s}$ and $x_{\rm c}$ are equally binned in logarithmic space in the

4. https://scikit-learn.org

DASH library (see Section A.2.2), we adopt $\log(c_{\rm h})$, $\log(c_{\rm s})$ and $\log(x_{\rm c})$ as the actual features. Furthermore, we train the model to predict $\log(f_{\rm b})$ as the target since $f_{\rm b}$ varies over several orders of magnitude. In the following analysis, only data points satisfying the numerical criteria of Paper II are included (a brief summary is given in Section A.2.3).

The ML algorithm we adopt, RF, is based on decision trees (Quinlan, 1986). While the decision tree method is intuitive and useful, trained models tend to be overfitted, i.e., the training data set is very accurately reproduced while poor predictions are made for untrained cases. In order to avoid this overfitting issue, RF constructs an ensemble of decision trees and adopts the mean prediction of individual decision trees as the final prediction. The ensemble consists of 20 decision trees with a maximum depth of 20. Here, the depth of the tree is the number of layers from a root to a leaf. The other hyperparameters are set to their default values in the `scikit-learn` implementation.

In order to increase the confidence in our trained model, we adopt group k-fold cross validation (e.g., Browne, 2000, and references therein). First, the full data set is divided into k subsets composed of data points from randomly selected simulations. Note that all time steps from each simulation are assigned to one group such that they are all placed within the same subset and thus are not split between multiple subsets. Then, the RF is trained with $k-1$ subsets and the trained model is tested on the remaining subset. This training and test procedure is iterated k times and the performance (R^2 score, see below for details) is measured as the average of the k models. We set $k = 5$ and the number of simulations in each subset is almost the same (435 or 436). To further verify the robustness of our trained model, we perform a test with a reduced number of $t/T_{\rm r}$ bins. While the other four parameters ($c_{\rm h}$, $c_{\rm s}$, $x_{\rm c}$, and η) have 11 bins for each, $t/T_{\rm r}$ has 301 bins in the DASH library. This might artificially weight $t/T_{\rm r}$ more strongly relative to the other features. To assuage this concern, we also train a model in which we use only 11 of the 301 epochs, equally spaced in time, such that all five features have 11 bins.

Fig. A.9 compares the evolution of $f_{\rm b}$ in the N-body simulations with the predictions made by the trained RF model and depicts that the model is predictive at the ~ 0.1 dex level. Note that the model predictions shown in Fig. A.9 are of simulations included only in the test fold for this particular trained model. In our case, a measure of accuracy of the trained model, the R^2 score (coefficient of determination) is defined as follows:

$$R^2 \equiv 1 - \frac{\sum[\log(f_{\rm b,sim}) - \log(f_{\rm b,pred})]^2}{\sum[\log(f_{\rm b,sim}) - \langle\log(f_{\rm b,sim})\rangle]^2}. \tag{A.17}$$

The summation runs over all data points in the test set. The mean value in the test set is simply given as

$$\langle\log(f_{\rm b,sim})\rangle = \frac{1}{N_{\rm test}}\sum\log(f_{\rm b,sim}). \tag{A.18}$$

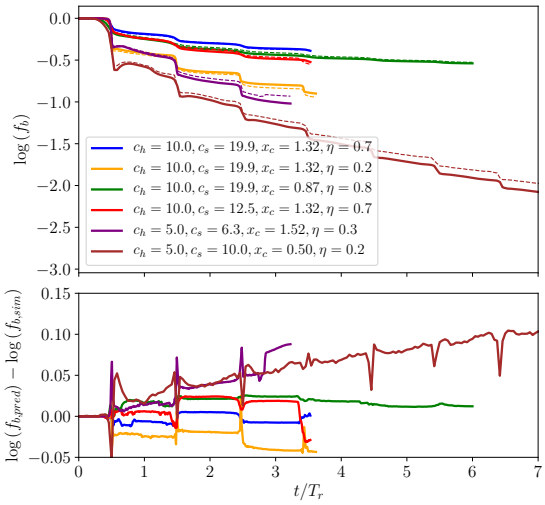

Figure A.9: Comparison between the subhalo bound mass fraction in the simulations, $f_{\rm b,sim}$, and those predicted by the model based on a commonly-used machine learning method, random forest regression, $f_{\rm b,pred}$. (*Upper*) Solid lines represent the simulation results and dashed lines represent the prediction by the model. (*Lower*) Residuals between the simulations and model predictions. The same colour scheme is used as the upper panel. The times are scaled by the radial period, $T_{\rm r}$.

Here, $N_{\rm test} = N_{\rm snap} \times N_{\rm sim}/5 \approx 131,000$ is the number of data points in the test set and $N_{\rm sim}(= 2,177)$ and $N_{\rm snap}(= 301)$ represent the number of simulations in the DASH library and the number of snapshots in each simulation, respectively. We emphasize again that only the data points fulfilling the numerical convergence criteria (see Paper II and Section A.2.1.4) are included in this analysis. The trained model yields $R^2 > 0.98$ for the test set in all five cross-validation cases, indicating that the trained model works well for untrained cases within the covered parameter space. The R^2 score does not change even if we reduce the number of $t/T_{\rm r}$ bins by a factor of 30, verifying that the higher number of bins in $t/T_{\rm r}$ does not matter for our model.

`scikit-learn` also reports the importance of features, i.e., how much the model depends on each feature for predicting $f_{\rm b}$. The derived importance for each feature is $[t/T_{\rm r}, \log(c_{\rm h}), \log(c_{\rm s}), \log(x_{\rm c}), \eta] = [0.46, 0.05, 0.21, 0.09, 0.19]$, meaning that $f_{\rm b}$ depends strongly on time while $c_{\rm s}$ and η play the most dominant roles among the four given parameters. This information may be useful in constructing semi-analytical models and in determining the parameter sets to be studied in the subsequent expansion of the DASH library.

The model trained with the full data set is saved in the file named `dash_fb_rf.joblib` and available within the Jupyter notebook, `dash_fb_predict.ipynb`, in the DASH library. Inputting the four parameters we vary, $x_{\rm c}$, η, $c_{\rm h}$, and $c_{\rm s}$, as well as the time of interest, one obtains the expected trajectories of the mass evolution of tidally stripped subhaloes. While the model is accurate and easy to use, it is not robust for extrapolation beyond the sampled region of parameter space because of the nature of decision tree-based

algorithms, which can only interpolate between the data points in the training set. The prediction made by RF corresponds to the mean prediction of individual decision trees and hence it does not work in the parameter space where no data points are found. Other ML algorithms, e.g., support vector machines and others based on neural networks, are needed to construct models that can be extrapolated, but these frameworks generally require more complicated data preprocessing treatment and tuning of hyperparameters than those for RF to obtain good models.

Subsequent studies will use the other types of data available in DASH, such as the radial profiles of mass density and velocity dispersion, in order to investigate the dynamical evolution of tidally stripped subhaloes in more detail (e.g., Green & van den Bosch, 2019).

A.5 Summary and discussion

Cosmological N-body simulations are the prime tool used to study the hierarchical assembly of dark matter haloes. They reveal that virialized dark matter haloes have a universal density profile (e.g., Navarro et al., 1997), and that roughly 10 percent of their mass is bound up in distinct subhaloes (e.g., Ghigna et al., 1998; Gao et al., 2004; Giocoli et al., 2010). According to the same simulations, a large fraction of these subhaloes completely disrupt after a few orbital periods (Han et al., 2016; van den Bosch, 2017). It has recently been argued that the majority of this disruption is artificial (Peñarrubia et al., 2010; van den Bosch et al., 2018; van den Bosch & Ogiya, 2018), and thus that state-of-the-art cosmological simulations still suffer from an appreciable amount of 'over-merging'. Most importantly, van den Bosch & Ogiya (2018) argued that this problem may go unnoticed in standard numerical 'convergence' tests. Hence, it is prudent that we consider alternative methods to predict the abundance and demographics of dark matter substructure, which is a potentially powerful Rosetta stone for deciphering the nature of dark matter.

One alternative to numerical simulations is a semi-analytical approach that combines halo merger trees, constructed using the framework of extended Press-Schechter theory (Bond et al., 1991), with a treatment of the tidal evolution of subhaloes as they orbit their host. These models are not hampered by discreteness issues or limiting force resolution responsible for artificial disruption. In addition, these models are far less CPU-intensive than actual N-body simulations, thus allowing for an extensive exploration of parameter space. Numerous models along this line have been constructed in the past (Taylor & Babul, 2001, 2004; van den Bosch et al., 2005b; Peñarrubia & Benson, 2005; Zentner et al., 2005a; Diemand et al., 2007; Kampakoglou & Benson, 2007; Gan et al., 2010; Pullen et al., 2014; Jiang & van den Bosch, 2016). Unfortunately, since we lack a purely analytical treatment of tidal stripping and heating, these models typically contain one or more 'fudge' parameters. These are tuned by requiring the model to reproduce the subhalo mass functions taken

from cosmological N-body simulations. The obvious downside of this approach is that the models thereby inherit the over-merging problems of the simulations.

In order to overcome this dilemma, we need idealized simulations that are (i) well resolved and free from artificial disruption, and (ii) optimized to allow for calibration of semi-analytical treatments of tidal stripping and heating. This chapter presents the DASH library, consisting of 2,177 idealized, high-resolution ($N = 1,048,576$), collisionless N-body simulations of individual dark matter subhaloes orbiting in the potential of a static, analytical host halo. The simulations have sufficient mass and force resolution to overcome artificial disruption (i.e., they satisfy the numerical reliability criteria given by equations [A.12] and [A.14]), and sample the entire parameter space of orbital energies, orbital angular momenta, and halo concentrations relevant for dark matter substructure. All simulations adopt a host halo-to-subhalo mass ratio of $\mathcal{M} = M_{\rm vir,h}/M_{\rm vir,s} = 1000$, for which dynamical friction, which is not accounted for in the DASH simulations, is negligible. Because of the scale-free nature of the tidal evolution of subhaloes (see Section A.3.1), the DASH simulations are valid for any mass ratio large enough such that dynamical friction is negligble (i.e., $\mathcal{M} \gtrsim 100$). Each simulation is evolved for a period of roughly 36 Gyr, during which the subhalo undergoes anywhere between 2 and 12 radial orbits. For each simulation, the DASH library, which is publicly available, contains simple text files that present, among others, the temporal evolution of the subhalo's bound mass fraction, and the density and velocity dispersion profiles of the bound particles of the subhalo at 301 outputs equally spaced in time. The library also contains a Python code, trained on the DASH simulation data, that uses random forest regression to predict the bound mass fraction of subhaloes as a function of time for given halo concentrations and orbital parameters. This code, which is accurate at the 0.1 dex level, conveniently summarizes the main results from our large suite of simulations.

In Green & van den Bosch (2019), we use the DASH library to calibrate a new and improved semi-analytical model of the tidal evolution of subhaloes, which we will subsequently use in combination with accurate halo merger trees (e.g., Parkinson et al., 2008; Jiang & van den Bosch, 2014) to predict the subhalo mass function of CDM haloes, unhindered by artificial disruption. This will shed new light on the level of reliability of the subhalo demographics that have been extracted from cosmological N-body simulations.

Finally, we emphasize that although the parameter space covered by the DASH library is vast, it is by no means exhaustive. One obvious shortcoming, as discussed above, is that the DASH simulations are inadequate to describe major mergers with $\mathcal{M} \lesssim 100$. In those cases, dynamical friction due to the host, and self-friction due to tidally stripped material, cause the orbit of the subhalo to decay, exposing it to stronger tides. Another degree of freedom not covered here is the inner density slope of dark matter haloes. It is well known that observations of dwarf galaxies often suggest that their haloes have constant density cores, rather than the steep r^{-1}-cusps predicted by dark matter-only simulations (e.g., Burkert, 1995; Gentile et al.,

2004; Oh et al., 2011; Hayashi & Chiba, 2015). Such cores can be created within the CDM paradigm by a variety of baryonic processes (e.g. El-Zant et al., 2001a; Inoue & Saitoh, 2011; Pontzen & Governato, 2012; Ogiya & Mori, 2014), and have a dramatic impact on the tidal evolution of subhaloes (Peñarrubia et al., 2010; Errani et al., 2015; Ogiya, 2018). In addition, baryons modify the potentials of host- and subhaloes through the bulges and discs that they form at the halo centres, and these also strongly impact the tidal fields (Errani et al., 2017; Garrison-Kimmel et al., 2017). Finally, in the DASH simulations presented here, the host halo is assumed to be spherically symmetric, which allows us to completely specify each orbit with only two parameters (energy and angular momentum). Cosmological simulations, though, indicate that dark matter haloes are expected to be triaxial systems (e.g., Jing & Suto, 2002; Allgood et al., 2006; Hayashi et al., 2007), consistent with the shapes of the gravitational potentials of galaxies and clusters as inferred from a variety of observations (e.g., Oguri et al., 2005; Corless & King, 2007; Law & Majewski, 2010). Triaxial systems have a much richer variety of orbits, which is likely to impact the tidal evolution of subhaloes.

In the near future, we therefore anticipate augmenting the DASH library with a suite of simulations that probe some of this extended parameter space. These additional simulations will be particularly useful for informing the semi-analytical treatments mentioned above. Ideally, any such semi-analytical treatment should capture the actual physics of tidal stripping and heating, and should thus be able to correctly predict the tidal evolution of subhaloes in triaxial potentials, in the presence of orbital decay due to dynamical friction, or in the case where the potential of the host- and/or subhalo has been modified due to the impact of baryonic processes. It remains to be seen to what extent the models can meet this challenge, and it is our hope that the DASH library presented here, as well as its future extensions, will play an important role in this process.

Appendix B

SatGen

This chapter has been published as an article by Jiang et al. (2021) in the Monthly Notices of the Royal Astronomical Society by Oxford University Press. Since I am a co-author of this work, I have elected to include it as an appendix because it provides a more thorough description of the `SatGen` model than that of Section 3.2.

B.1 Background

In our modern understanding of the Universe, structures form hierarchically: dark matter (DM) overdensities collapse into gravitationally bound haloes, which merge to form larger haloes. The smaller participant of a merger survive as substructure within the merger remnant, experiencing tidal interactions, losing mass, and undergoing structural change. Galaxies form inside DM haloes. When a halo merger occurs, the less massive progenitor becomes a substructure and the inhabiting galaxy becomes a satellite galaxy. Subhaloes and satellites are therefore the building blocks of host haloes and central galaxies and serve as relics of structures that formed earlier, with their demographics containing the information of the assembly history of the host system as well as the Universe at large.

Apart from their cosmological significance, satellite galaxies are interesting on their own, in the sense that galaxies of extreme morphology are usually spotted in dense environments. For example, among bright dwarfs (i.e., galaxies with stellar mass $m_\star \sim 10^{7-9}\mathrm{M}_\odot$) in the Local Group or in galaxy clusters, galaxies range from ultra-compact dwarfs (UCDs, with half-stellar-mass radii $l_{\rm eff} \sim 0.1$ kpc, e.g., Drinkwater et al. 2003) to ultra-diffuse galaxies (UDGs, with $l_{\rm eff} \sim 5$ kpc, e.g., van Dokkum et al. 2015), spanning almost 2 dex in size. The environment may be the key to such diversity: the central galaxy and the host halo can make a satellite

more diffuse or more compact through tidal effects depending on the initial conditions, the time since the infall of the satellite, and the orbit of the satellite.

Subhaloes and satellites have been studied using numerical simulations (e.g. Gao et al., 2004; Diemand et al., 2008; Springel et al., 2008; Wu et al., 2013; Garrison-Kimmel et al., 2014a; Mao et al., 2015; Sawala et al., 2016; Wetzel et al., 2016; Garrison-Kimmel et al., 2019) and semi-analytical models (e.g. Taylor & Babul, 2001; Benson et al., 2002a,b; Zentner & Bullock, 2003; Zentner et al., 2005a; Gan et al., 2010; Jiang & van den Bosch, 2016; Nadler et al., 2019; Yang et al., 2020). Cosmological N-body simulations produce a plethora of subhaloes compared to observed satellite galaxies. While low-mass haloes ($M_{\rm vir} \lesssim 10^9$ M$_\odot$) are expected to be truly dark due to the suppression of star formation by the cosmic UV background, thereby alleviating this "missing satellite" problem (e.g. Benson et al., 2002a,b; Hambrick et al., 2011), a more persistent challenge lies in the overabundance of massive and dense subhaloes – they are too big to fail forming stars (Boylan-Kolchin et al., 2011). The "too-big-to-fail" problem is not merely the overabundance of massive satellites, but also highlights the lack of structural diversity in the simulated satellite populations (e.g. Jiang & van den Bosch, 2015) – the simulated population of massive satellites are dense in their centres, showing a narrow distribution of maximum circular velocities ($v_{\rm max}$), while the observed bright dwarf satellites exhibit a larger variety of inner densities (Oman et al., 2015) and a broad distribution of $v_{\rm max}$. Hydro-simulations have shown that including baryons can help to reduce the abundance of massive satellites, mostly because the central galaxies enhance the tidal disruption of satellites (e.g., Garrison-Kimmel et al. 2019, but see also Errani et al. 2017 and Garrison-Kimmel et al. 2017, which use idealized N-body simulations with a galactic disc). However, hydro-simulations still do not fully reproduce the structural diversity of dwarf satellites (e.g. Garrison-Kimmel et al., 2019), missing the most diffuse and most compact dwarf satellites seen around the Milky Way (MW) and M31.

The limitations of cosmological simulations can be summarized as follows. First, simulating a satellite population is computationally expensive – it requires a large dynamical range in mass and in spatial scale. State-of-the-art zoom-in simulations typically produce on the order ~10 MW-like host systems (e.g. Sawala et al., 2016; Garrison-Kimmel et al., 2019) or ~1 cluster (e.g. Pillepich et al., 2019; Tremmel et al., 2019), whereas quantifying the cosmic variance of satellite statistics for a given host mass requires at least hundreds of random realizations (Purcell & Zentner, 2012; Jiang & van den Bosch, 2015). Second, artificial disruption of satellites due to insufficient resolution is still prevalent in modern simulations. It is alarming to realize that, in the Bolshoi simulation (Klypin et al., 2011), ~60% of subhaloes with infall mass larger than 10% of the instantaneous host halo mass cannot even survive for one orbit (Jiang & van den Bosch, 2017) and ~13% of subhaloes are disrupted per Gyr (van den Bosch, 2017), despite the use of a sophisticated, phase-space based halo finder (Behroozi et al., 2013a). Similar results have been reported for zoom-in sim-

ulations: about half of the subhaloes in the Aquarius simulations have been disrupted, irrespective of their masses at infall (Han et al., 2016). Idealized simulations (of higher resolution than cosmological ones) reveal that satellite disruption is mostly numerical in origin, caused mainly due to inadequate force softening and a runaway instability triggered by the amplification of discreteness noise in the presence of a tidal field (van den Bosch et al., 2018; van den Bosch & Ogiya, 2018). Third, halo finding algorithms, especially those based only on identifying instantaneous overdensities, have difficulty in recovering subhaloes when they are located in dense region of the host (Muldrew et al., 2011; van den Bosch & Jiang, 2016).

Semi-analytical models serve as complementary tools to simulations in the study of satellite galaxies and outperform simulations in terms of statistical power and numerical resolution. Such models consist of halo merger trees and analytical prescriptions for satellite evolution. Most of these models focus on the DM components, using cuspy profiles (Navarro et al., 1997) to describe both the host halo and the satellites, ignoring baryonic components and processes. However, hydro-simulations have shown that baryonic influence cannot be neglected for satellites. First, the DM profile of satellites at infall is not necessarily cuspy. For example, supernovae-driven gas outflows can create dark matter cores (e.g., Pontzen & Governato, 2012) and systems with cored profiles follow different tidal evolution paths than cuspy ones with the same initial orbit (e.g., Peñarrubia et al., 2010). Second, the central galaxy, e.g., a MW-like disc, can significantly impact the spatial distribution of a satellite population by reducing the survivability of the satellites that travel across the disc-dominated region (e.g., Garrison-Kimmel et al., 2017). Finally and obviously, to study the baryonic properties of satellite galaxies instead of merely the statistics of DM subhaloes, the baryonic components of a satellite and their evolution in a dense environment must be considered. Hence, semi-analytical models of satellites are urgently in the need of upgrades in order to catch up with recent developments in cosmological simulations.

In this chapter, we present `SatGen` , a new semi-analytical model for generating merger trees and evolving satellite populations, and then, as a proof-of-concept for `SatGen` , we perform a study of satellite statistics for MW/M31-like hosts. Compared to previous models, `SatGen` improves on several important aspects. First, it considers baryonic effects, both within the satellites and the host galaxy, on the structure and survivability of subhaloes. Subhaloes in `SatGen` can be described by profiles that have the flexibility to capture DM cores and that have been widely used to describe subhaloes in simulations, including a subclass of the $\alpha\beta\gamma$ profiles (Zhao, 1996; Dekel et al., 2017; Freundlich et al., 2020b) and the Einasto (1965) profile. The initial structure of the subhaloes are based upon halo response models extracted from state-of-the-art hydro-simulations and analytical modeling; by changing the halo response model, the user can make `SatGen` emulate different simulations. Host systems in `SatGen` can be composed of (a combination of) a baryonic disc, stellar bulge, and DM halo. Second, `SatGen` incorporates simple recipes for the evolution of

the stellar and gaseous components of satellite galaxies. The structural evolution recipes of subhaloes and stellar components are either analytical and physically motivated or extracted from high-resolution idealized simulations, which makes `SatGen` essentially free from the effects of numerical disruption of satellites commonly seen in cosmological simulations. Finally, in keeping with the most sophisticated previous models of this kind (e.g., Taylor & Babul, 2001; Benson et al., 2002a; Zentner et al., 2005a), `SatGen` follows the orbit of each satellite, while accounting for dynamical friction.

This chapter is organized as follows. In Section B.2, we describe the model. In Section B.4, we present satellite statistics of MW/M31-sized systems, comparing model predictions with observations (Section B.4.1), as well as comparing model results using different halo response models characteristic of different hydro-simulations (Section B.4.2). We also quantify the effect of a baryonic disc potential on the abundance, spatial distribution, and internal structure of satellites (Section B.4.3). In Section B.5, we explore the conditions for a massive satellite to survive (or get disrupted) in a MW/M31 potential. In Section B.6, we summarize the model and our findings.

Throughout, we use m and M to indicate satellite mass and host mass, respectively. We use l and r to refer to satellite-centric radius and host-centric distance, respectively. Thus, a density profile written as $\rho(r)$ refers to that of the host system and written as $\rho(l)$ refers to that of the satellite. We define the virial radius of a distinct halo as the radius within which the average density is $\Delta = 200$ times the critical density for closure. We adopt a flat ΛCDM cosmology with the present-day matter density $\Omega_{\rm m} = 0.3$, baryonic density $\Omega_{\rm b} = 0.0465$, dark energy density $\Omega_\Lambda = 0.7$, a power spectrum normalization $\sigma_8 = 0.8$, a power-law spectral index of $n_s = 1$, and a Hubble parameter of $h = 0.7$. All of these assumptions can be changed easily in `SatGen`.

B.2 Model

The model builds upon halo merger trees. Combining these merger trees with some empirical prescriptions from simulations, we obtain the initial masses, profiles, and baryonic properties of satellites. Then, we follow the orbits of the satellites, modeling tidal stripping and the structural evolution of both the DM and baryonic components. The `SatGen` code is made publicly available on GitHub.[1] A schematic view of the model is presented in Fig. B.1. Below, we introduce each model component in sufficient detail to reproduce the exercise in Section B.4, leaving more comprehensive details in the appendices. Readers who want to see the results first with a basic idea of how the model works can view Fig. B.1 and read Section B.2.8 for a quicker overview and jump to Section B.4.

1. https://github.com/shergreen/SatGen

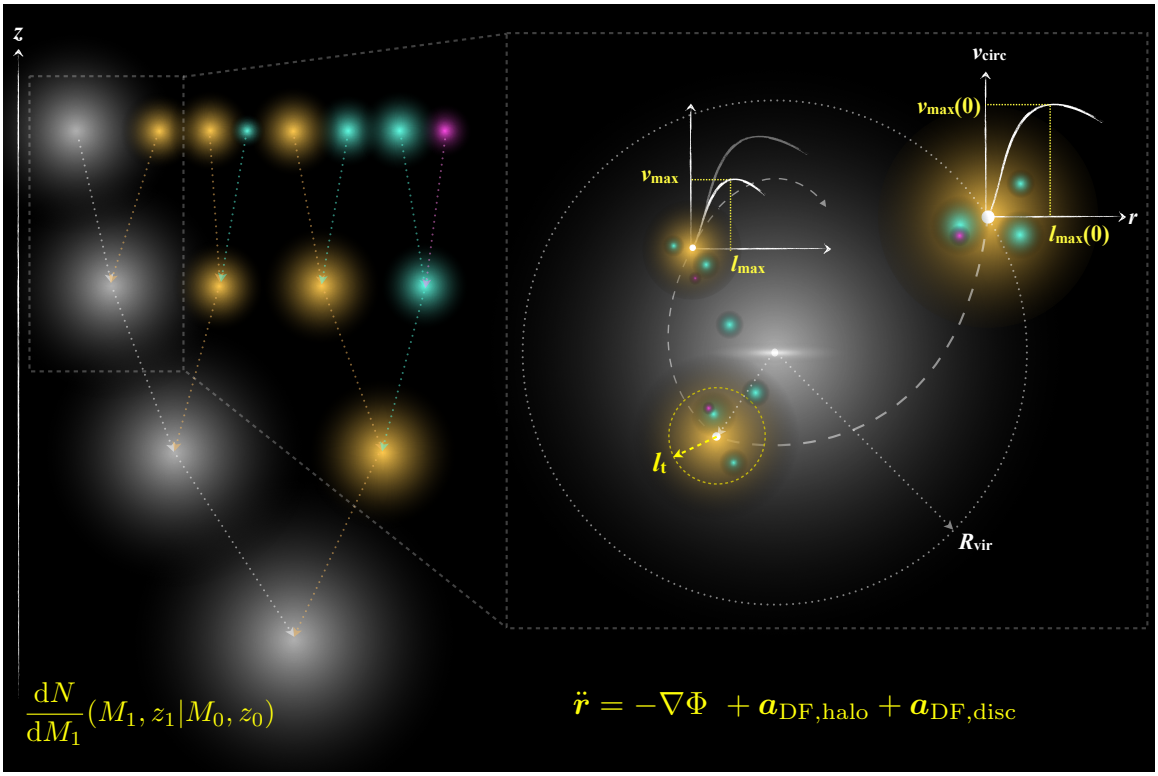

Figure B.1: Schematic view of the `SatGen` model. *Left*: a halo merger tree, generated by sampling the EPS progenitor mass function, $\mathrm{d}N/\mathrm{d}M_1(M_1, z_1|M_0, z_0)$ (see Section B.2.1). Different colours differentiate branches of different levels – the main branch (i.e., the host-halo branch) is white; the branches of the first-order satellites, i.e., the satellites that are directly accreted by the host system, are yellow; the branches of the second-order satellites, i.e., the satellites that directly merge with first-order satellite progenitors and are brought into the host halo as sub-substructures, are cyan; and so on. *Right*: a zoom-in view of what happens after a satellite is accreted. In this illustration, a first-order satellite orbits around a host composed of a smooth halo and a galactic disc (see Section B.2.3 for how we initialize the host). The satellite brings its own higher-order substructure to the host, loses mass (see Section B.2.5 for how we model tidal stripping), releases higher-order satellites, and evolves in structure (represented by the peak circular velocity, $v_{\rm max}$, and the corresponding location, $l_{\rm max}$), as illustrated by the schematic plots of the circular velocity profiles at infall (see Section B.2.3 for how we initialize subhalo structure at infall) and at a later epoch when it is significantly stripped (see Section B.2.6 for how the structural evolution is modeled). For such an eccentric orbit (see Section B.2.3 for how we draw initial orbits), tidal stripping is most efficient at the orbital pericentre, where the Hill surface is indicated by a yellow dotted circle and the tidal radius, $l_{\rm t}$, is marked (see Section B.2.5 for how we model tidal stripping). For such a major merger, orbital decay due to dynamical friction (Section B.2.4) is significant, as illustrated by the dashed line. Not shown here are the prescriptions for the initialization and the evolution of the stellar and gaseous components of the satellite (see Section B.2.3 and Section B.2.6 for details).

B.2.1 Halo merger trees

`SatGen` generates halo merger trees using an algorithm (Parkinson et al., 2008) based on the extended Press-Schechter (EPS) formalism (Lacey & Cole, 1993). The EPS method provides the expected number of progenitor haloes of mass M_1 at redshift z_1 for a target halo of mass M_0 at redshift $z_0 < z_1$,

$$\frac{\mathrm{d}N}{\mathrm{d}M_1}(M_1, z_1|M_0, z_0)\mathrm{d}M_1 = \frac{M_0}{M_1}\frac{\Delta\omega}{\sqrt{2\pi}(\Delta S)^{3/2}}e^{-\frac{(\Delta\omega)^2}{2\Delta S}}\left|\frac{\mathrm{d}S}{\mathrm{d}M}\right|_{M_1}\mathrm{d}M_1, \tag{B.1}$$

where $S \equiv \sigma^2(M)$ is the variance of the density field linearly extrapolated to $z = 0$ and smoothed with a sharp k-space filter of mass M, $\delta(z)$ is the critical overdensity for spherical collapse, $\Delta S = \sigma^2(M_1) - \sigma^2(M_0)$, and $\Delta\omega = \delta_\mathrm{c}(z_1) - \delta_\mathrm{c}(z_0)$.

However, it has been shown that merger trees constructed by strictly sampling this progenitor mass distribution over-predict the low-redshift merger rate compared to cosmological simulations (e.g., Zhang et al., 2008; Jiang & van den Bosch, 2014). In order to achieve better agreement with simulations, we follow Parkinson et al. (2008) by adding a corrective factor of the following form to the right-hand side of equation (B.1):

$$G(M_1|M_0, z_0) = G_0\left(\frac{S_1}{S_0}\right)^{\frac{\gamma_1}{2}}\left(\frac{\omega_0^2}{S_0}\right)^{\frac{\gamma_2}{2}}, \tag{B.2}$$

where $S_1 = \sigma^2(M_1)$, $S_0 = \sigma^2(M_0)$, $\omega_0 = \delta^2(z_0)$, and we adopt $G_0 = 0.6353$, $\gamma_1 = 0.1761$, and $\gamma_2 = 0.0411$ following Benson (2017).

We construct merger trees using the time-stepping advocated in Appendix A of Parkinson et al. (2008), which corresponds to $\Delta z{\sim}0.001$.

In order to reduce memory usage, we follow Jiang & van den Bosch (2016) and down-sample the temporal resolution of the trees by only registering progenitor haloes every timestep of $\Delta t = 0.1 t_\mathrm{dyn}(z)$, where $t_\mathrm{dyn} = \sqrt{3\pi/[16G\Delta\rho_\mathrm{crit}(z)]}$ is the instantaneous dynamical time of DM haloes.

B.2.2 Profiles for DM haloes and baryonic discs

With `SatGen`, one has multiple choices for the profile of a DM halo, including the Navarro et al. (1997, hereafter NFW) profile, the Einasto (1965, hereafter Einasto) profile, and the Dekel et al. (2017, hereafter Dekel+) profile, which is a subclass of the $\alpha\beta\gamma$ profiles (Zhao, 1996). Galactic discs and bulges can be described by the Miyamoto & Nagai (1975, hereafter MN) profile and the Einasto profile, respectively. One can set up a host system using a combination of the aforementioned profiles, e.g., a NFW halo plus an embedded MN disc. In Appendix B.3.1, we provide analytical expressions for the profiles of density, enclosed mass, gravitational potential, and velocity dispersion of all of the supported profiles. Here, we describe the Dekel+ halo profile

and the MN profile, which will be used in the experiments in Section B.4.

B.2.2.1 Dekel et al. (2017) halo profile

A Dekel+ halo is defined by four parameters: the virial mass, $M_{\rm vir}$, a concentration parameter, c, the (negative of the) logarithmic density slope in the centre, $\alpha = -{\rm d}\ln\rho/{\rm d}\ln r|_{r\to 0}$, and the spherical overdensity, Δ. The density profile is given by:

$$\rho(r) = \frac{\rho_0}{x^\alpha(1+x^{1/\beta})^{\beta(\gamma-\alpha)}}, \quad \beta = 2, \quad \gamma = 3+\beta^{-1} = 3.5, \tag{B.3}$$

where $x \equiv r/r_{\rm s}$ is the radius scaled by an intermediate radius $r_{\rm s}$ that is related to $r_{\rm vir}$ by the concentration parameter, $r_{\rm s} = r_{\rm vir}/c$, and $\rho_0 = [c^3(3-\alpha)/3/f(c,\alpha)]\Delta\rho_{\rm crit}$, with $f(x,\alpha) = \chi^{2(3-\alpha)}$ and $\chi \equiv x^{1/2}/(1+x^{1/2})$.

The Dekel+ profile has only one more degree of freedom than the NFW profile and it has three merits that make it ideal for use in semi-analytical models. First, it can accurately describe haloes in hydro-simulations (Dekel et al., 2017; Freundlich et al., 2020b), having enough flexibility near the centre to accurately describe the cusp-core transformation (Freundlich et al., 2020a). Second, it has an outer slope of $\gamma = 3.5$, steeper than that of the NFW profile and thus more appropriate for describing subhaloes that are stripped. Finally, it has fully analytical expressions for the profiles of enclosed mass, gravitational potential, and velocity dispersion, facilitating fast orbit integration and making it more convenient to use than the Einasto profile or other subclasses of the $\alpha\beta\gamma$ family (see more details in Freundlich et al., 2020b). [2] The mass inside radius r is given by

$$M(r) = M_{\rm vir}\frac{f(x,\alpha)}{f(c,\alpha)}, \tag{B.4}$$

the gravitational potential can be expressed as

$$\Phi(r) = -V_{\rm vir}^2\frac{2c}{f(c,\alpha)}\left[\frac{1-\chi^{2(2-\alpha)}}{2(2-\alpha)} - \frac{1-\chi^{2(2-\alpha)+1}}{2(2-\alpha)+1}\right], \tag{B.5}$$

where $V_{\rm vir}$ is the virial velocity, and the one-dimensional isotropic velocity dispersion $\sigma(r)$ is given by

$$\sigma^2(r) = 2V_{\rm vir}^2\frac{c}{f(c,\alpha)}\frac{x^{3.5}}{\chi^{2(3.5-\alpha)}}\sum_{i=0}^{8}\frac{(-1)^i 8!}{i!(8-i)!}\frac{1-\chi^{4(1-\alpha)+i}}{4(1-\alpha)+i}. \tag{B.6}$$

Unlike the NFW profile, where the scale radius $r_{\rm s}$ is the same as the radius at which the logarithmic density slope equals -2 (hereafter referred to as r_2), in a Dekel+ profile, the two radii are related by $r_2 =$

2. In fact, a full family of profiles of the form of equation (B.3) with $\beta = n$ and $\gamma = 3 + k/n$ (where k and n are integers) have fully analytical expressions for the profiles of potential and velocity dispersion (Zhao, 1996). The choice of $n = 2$ and $k = 1$, as in the Dekel+ profile, yields accurate enough descriptions of haloes in hydro-simulations.

$[(2-\alpha)/1.5]^2 r_s$. That is, the conventional concentration parameter, $c_2 = r_{vir}/r_2$, is related to the Dekel+ concentration by

$$c_2 = \left(\frac{1.5}{2-\alpha}\right)^2 c. \tag{B.7}$$

The radius of peak circular velocity, r_{max}, is related to r_2 by

$$r_{max} = 2.25 r_2 = (2-\alpha)^2 r_s. \tag{B.8}$$

The parameter α is the logarithmic density slope, $-\mathrm{d}\ln\rho/\mathrm{d}\ln r$, in the asymptotic limit $r \to 0$, which may fall well outside the radial range of interest (for example between $0.01 r_{vir}$ and r_{vir}). For the slope in the radial range of interest, the slope profile is given by

$$s(r) = -\frac{\mathrm{d}\ln\rho}{\mathrm{d}\ln r} = \frac{\alpha + 3.5\sqrt{x}}{1+\sqrt{x}}. \tag{B.9}$$

The slope at $0.01 r_{vir}$, widely used in the context of the cusp-core issue, is

$$s_{0.01} \equiv s(0.01 r_{vir}) = \frac{\alpha + 0.35\sqrt{c}}{1+0.1\sqrt{c}}. \tag{B.10}$$

For $s_{0.01}$ values that are commonly seen in simulations and observations ($0-2$) and for a typical concentration (e.g., $c = 10$), we have $\alpha \in (-1.11, 1.53)$. That is, α can be negative for realistic profiles (corresponding to a density that actually decreases towards the halo centre) and thus $s_{0.01}$ is a more physical quantity than α when it comes to comparing the cuspiness of density profiles.

B.2.2.2 Miyamoto & Nagai (1975) disc profile

A MN disc is specified by three parameters: the disc mass (M_d), a scale radius (a), and a scale height (b). The density and potential profiles are given by

$$\rho(R,z) = \frac{M_d b^2}{4\pi} \frac{aR^2 + (a+3\zeta)(a+\zeta)^2}{\zeta^3 [R^2 + (a+\zeta)^2]^{5/2}} \tag{B.11}$$

and

$$\Phi(R,z) = -\frac{GM_d}{\sqrt{R^2 + (a+\zeta)^2}}, \tag{B.12}$$

respectively, where $\zeta = \sqrt{z^2+b^2}$ and and R, ϕ, and z are the cylindrical coordinates. For an axisymmetric disc whose distribution function only depends on E and L_z, the radial and axial velocity dispersions are equal: $\sigma_R = \sigma_z \equiv \sigma$. Further assuming that the disc is an isotropic rotator, i.e., $\overline{V}_\phi^2/(\overline{V_\phi^2} - \sigma^2) = 1$, we have

$\sigma_\phi^2 = \overline{V_\phi^2} - \overline{V}_\phi^2 = \sigma^2$, and σ^2 is given by Ciotti & Pellegrini (1996) by

$$\sigma^2(R,z) = \frac{GM_{\rm d}^2 b^2}{8\pi\rho(R,z)} \frac{(a+\zeta)^2}{\zeta^2[R^2+(a+\zeta)^2]^3}. \tag{B.13}$$

The net rotation, $\overline{V}_\phi$, can therefore be expressed by

$$\overline{V}_\phi^2 = V_{\rm circ}^2 + \frac{R}{\rho}\frac{\partial(\rho\sigma^2)}{\partial R} = \frac{GM_{\rm d}^2 ab^2}{4\pi\rho} \frac{R^2}{\zeta^3[R^2+(a+\zeta)^2]^3}, \tag{B.14}$$

where $V_{\rm circ}^2(R,z) = R\partial\Phi/\partial R$ and $(R/\rho)\partial(\rho\sigma^2)/\partial R$ is the asymmetric-drift term. Equations (B.13) and (B.14) are useful for modeling dynamical friction (Section B.2.4).

B.2.3 Initial conditions for satellite galaxies

The initial conditions for a satellite galaxy include (1) the properties of the host system when the satellite enters the virial sphere, (2) the orbit of the incoming satellite, and (3) the DM, stellar, and gaseous properties of the incoming satellite. Here we describe them one by one.

B.2.3.1 Initial host profile

The host halo mass is known from the main branch (i.e., the branch that tracks the most massive progenitor) of the merger tree. To fully specify the host halo profile, we also need the structural parameter(s). The halo concentration can be obtained from an empirical relation calibrated via simulations (Zhao et al., 2009), which relates the main branch merging history to the concentration parameter, c_2, by

$$c_2(M_{\rm vir},z) = \left\{4^8 + \left[\frac{t(z)}{t_{0.04}(M_{\rm vir},z)}\right]^{8.4}\right\}^{1/8}, \tag{B.15}$$

where $t(z)$ is the cosmic time at redshift z and $t_{0.04}$ is the cosmic time when the host halo has assembled 4% of its instantaneous mass, $M_{\rm vir}(z)$, which we extract from the halo's merger tree as described in Section B.2.1. If the host system is only an NFW halo, then concentration and mass completely specifies it. For a more complicated setup, e.g., a Dekel+ halo with an embedded MN disc, one needs additional assumptions depending on the system of interest (see e.g., Section B.4 for more details for MW/M31 analogues). The concentration c and the slope α of a Dekel+ halo can be obtained from equations (B.7), (B.10), and (B.15), with an assumption for $s_{0.01}$ that will be described in Section B.2.3.3.

B.2.3.2 Initial orbit

The initial orbit of a satellite can be specified by four pieces of information – the location of virial-crossing, orientation of the orbital plane, orbital energy, and orbital circularity. We assume that the infall locations are isotropically distributed on the virial sphere, and thus randomly draw an azimuthal angle (ϕ) from $[0, 2\pi]$ and a cosine polar angle ($\cos\theta$) from $[0, 1]$. We parameterize the specific energy of an orbit, E, by a unitless parameter, $x_{\rm c} = r_{\rm circ}(E)/r_{\rm vir}$, which is the radius of the circular orbit corresponding to the same orbital energy, E, in units of the virial radius of the host halo (e.g., van den Bosch, 2017). Orbital circularity, $\epsilon = j/j_{\rm circ}(E)$, is the ratio between the specific orbital angular momentum and that of a circular orbit of the same orbital energy. We assume $x_{\rm c} = 1$, typical of cosmological orbits seen in simulations[3] and draw ϵ from a distribution, $\mathrm{d}P/\mathrm{d}\epsilon = \pi\sin(\pi\epsilon)/2$, which approximates the ϵ distribution of infalling satellites measured in cosmological simulations (e.g., Wetzel, 2011; Jiang et al., 2015; van den Bosch, 2017).

For orbit integration (Section B.2.4), we need to translate these orbital parameters $(\phi, \theta, x_{\rm c}, \epsilon)$ to the position vector, $\boldsymbol{r}$, and the velocity vector, $\boldsymbol{V}$. Since `SatGen` supports axisymmetric potentials, we work in the cylindrical coordinate system, i.e., $\boldsymbol{r} = (R, \phi, z)$ and $\boldsymbol{V} = (V_R, V_\phi, V_z)$. The initial speed at virial-crossing (V) is given by

$$V = \sqrt{2[\Phi(x_{\rm c} r_{\rm vir}) - \Phi(r_{\rm vir})] + V^2_{\rm circ}(x_{\rm c} r_{\rm vir})}, \tag{B.16}$$

which is simply $V_{\rm vir}$ for $x_{\rm c} = 1$. Using the definition of ϵ, we can derive the angle ($\tilde{\theta}$) between $\boldsymbol{V}$ and $\boldsymbol{r}$:

$$\tilde{\theta} = \pi - \arcsin\left(\epsilon x_{\rm c} \frac{V_{\rm vir}}{V}\right). \tag{B.17}$$

In order to fully specify the orientation of the orbital plane, we need another angle for the velocity vector. We choose this angle to be the azimuthal angle ($\tilde{\phi}$) of $\boldsymbol{V}$ in the $\hat{\boldsymbol{\theta}}$-$\hat{\boldsymbol{\phi}}$-$\hat{\boldsymbol{r}}$ frame, and draw $\tilde{\phi}$ randomly from

3. To be more accurate, one can draw $x_{\rm c}$ from orbital energy distributions extracted from simulations (e.g., van den Bosch, 2017), which show a median value around $x_{\rm c}$~1. We opt to keep it simple and use $x_{\rm c} = 1$ in this work. After all, the correlation between initial orbital parameters and initial satellite properties is not clear yet. In Green et al. (2021a), we expand `SatGen` to draw orbits according to a distribution extracted from cosmological simulations, following Li et al. (2020).

$[0, 2\pi]$. Finally, we can specify all the phase-space coordinates of the infalling satellite:

$$\begin{aligned} R &= r_{\rm vir} \sin\theta, \\ \phi &= \phi, \\ z &= r_{\rm vir} \cos\theta, \\ V_R &= V(\sin\tilde{\theta}\cos\tilde{\phi}\cos\theta + \cos\tilde{\theta}\sin\theta), \\ V_\phi &= V\sin\tilde{\theta}\sin\tilde{\phi}, \\ V_z &= V(\cos\tilde{\theta}\cos\theta - \sin\tilde{\theta}\cos\tilde{\phi}\sin\theta). \end{aligned} \tag{B.18}$$

B.2.3.3 Initial subhalo density profiles

In cosmological N-body simulations, halo density profiles are well-approximated by NFW profiles. Therefore, if `SatGen` is used to emulate an N-body simulation, in order to initialize a subhalo profile we only need to compute the concentration parameter c_2 using equation (B.15).

To emulate hydro-simulations, we need to account for the fact that haloes react to baryonic processes that cause their profiles to deviate from NFW. The *halo response* to baryonic processes is mass-dependent (e.g., Di Cintio et al., 2014a; Dutton et al., 2016; Tollet et al., 2016; Freundlich et al., 2020a): qualitatively, low-mass haloes ($\lesssim 10^{11}$ $\mathrm{M}_\odot$) are susceptible to supernovae-driven gas outflows, becoming less concentrated and developing a flatter core; in contrast, massive haloes ($> 10^{12}$ $\mathrm{M}_\odot$) tend to contract as cold gas condenses in the centre, becoming cuspier. The halo response strength depends on the sub-grid physics adopted in the simulations. This is especially relevant for massive dwarf galaxies ($M_{\rm vir} \sim 10^{10.5}$ $\mathrm{M}_\odot$). Notably, simulations featuring bursty star formation, and thus strong episodic supernovae outflows, yield a strong halo response, whereas simulations with smooth, continuous star formation exhibit a negligible halo response in the dwarf regime (Bose et al., 2019; Dutton et al., 2019). The nature of the star formation burstiness, and thus the strength of the halo response, is closely related to the sub-grid recipe for star formation and is still highly uncertain and under debate.

Following ~~by~~ Di Cintio et al. (2014a,b) and Tollet et al. (2016), we parameterize the halo response with two relations: (1) the ratio of the hydro-simulation concentration and the corresponding DM-only concentration, $c_2/c_{2,\rm DMO}$ as a function of the stellar-to-halo-mass ratio (SHMR), $X = M_\star/M_{\rm vir}$, and (2) the logarithmic DM density slope measured at ~1% of the virial radius, $s_{0.01}$, as a function of the SHMR. Specifically, the concentration ratio can be expressed by

$$\frac{c_2}{c_{2,\rm DMO}} = a_0 + a_1 X^{b_1} - a_2 X^{b_2}, \tag{B.19}$$

where the constants a_i and b_i are simulation-specific and are chosen according to the simulation that one wishes `SatGen` to emulate. For example, we find that $(a_0, a_1, a_2) = (1.14, 186, 1)$ and $(b_1, b_2) = (1.37, 0.142)$ describe the halo response of the NIHAO (Wang et al., 2015) simulations accurately (Freundlich et al., 2020b). For these parameters, $c_2/c_{2,\rm DMO}$ approaches unity at $M_\star/M_{\rm vir} < 10^{-4}$, where star formation is weak and feedback effects are minimal (typical of low-mass haloes), is less than unity (~ 0.7) at $M_\star/M_{\rm vir} \sim 10^{-2.5}$ (typical of massive dwarf galaxies where feedback effects are maximal), and becomes > 1 at $M_\star/M_{\rm vir} > 10^{-2}$ (where adiabatic contraction dominates). Similarly, the inner density slope $s_{0.01}$ can be expressed as

$$s_{0.01} \equiv -\frac{{\rm d}\ln\rho}{{\rm d}\ln r}|_{0.01r_{\rm vir}} = \log\left[n_1\left(1+\frac{X}{X_1}\right)^{-\xi_1} + \left(\frac{X}{X_0}\right)^{\xi_0}\right] + n_0, \tag{B.20}$$

where the constants X_i, n_i, and ξ_i are, again, chosen to reflect the simulation sub-grid physics of interest (Tollet et al., 2016). For the NIHAO simulations, Freundlich et al. (2020b) find that $(n_0, n_1) = (1.45, 1)$, $(\xi_1, \xi_0) = (2.14, 0.21)$, and $(X_0, X_1) = (2.54\times10^{-3}, 9.87\times10^{-4})$. This describes the phenomenon that DM cores form if $X \sim 10^{-3}$-10^{-2}, cusps remain present for smaller X, and baryons deepen the gravitational potential at larger X. We add random Gaussian noise with $\sigma = 0.1$ and 0.18 to the $c_2/c_{2,\rm DMO}$ and $s_{0.01}$ values, respectively, based on Freundlich et al. (2020b) and Tollet et al. (2016). We note that the aforementioned halo response is likely quite generic for simulations featuring bursty star formation and episodic strong feedback, such as the FIRE simulations (Hopkins et al., 2014, 2018).

We use the Dekel+ profile to describe subhaloes affected by feedback. From equation (B.3), we can show that the slope at $r \to 0$ (α) and the slope at $r = 0.01r_{\rm vir}$ ($s_{0.01}$) are related by

$$\alpha = s_{0.01}(1 + 0.1\sqrt{c}) - 0.35\sqrt{c}. \tag{B.21}$$

Using equations (B.7), (B.15), (B.20), (B.21), and a SHMR, we can completely specify a Dekel+ subhalo at infall. [4]

We emphasize that one of the goals of `SatGen` is to quantify the influence of different halo response models on satellite statistics, and thus to distinguish the underlying sub-grid recipes adopted in simulations using observed satellite statistics. More specifically, the logic is the following. On the theory side, while it is computationally expensive to run simulations with adequate resolution for studying satellite galaxies, it is relatively cheap to simulate a suite of field galaxies that cover a wide range in mass and SHMR. These types of simulation suites, e.g., FIRE/FIRE-II (Hopkins et al., 2014, 2018), NIHAO (Wang et al., 2015),

4. For Einasto profiles, an expression analogous to equation (B.21) between the Einasto shape index and $s_{0.01}$ can be derived. See Appendix B.3.1 for details.

APOSTLE (Sawala et al., 2016), and Auriga (Grand et al., 2017), provide us with halo response templates, $(c_2/c_{2,\mathrm{DMO}})(X)$ and $s_{0.01}(X)$ (e.g., Tollet et al., 2016; Bose et al., 2019), which are used as inputs for the `SatGen` model. `SatGen` then propagates the difference in halo response to satellite structures because, as will be detailed in Section B.2.6, satellites of different initial structures evolve differently in response to tidal effects. In this way, `SatGen` produces satellites as would be produced by high-resolution simulations using the corresponding sub-grid recipe. On the observational side, galaxy structure and halo structure measurements are usually performed on galaxies of known distances, which are typically satellites. By propagating the baryonic effects obtained from zoom-in simulations of centrals onto satellite populations, `SatGen` facilitates the comparison between theory and observation.

B.2.3.4 Initial baryonic properties

Apart from subhalo properties, we also model the stellar mass, stellar size, and gas distribution. We assign a stellar mass to an infalling satellite using the SHMR from halo abundance matching. In particular, we use the expression of stellar mass ($M_\star$) as a function of halo mass (M_{vir}) and redshift z by Rodríguez-Puebla et al. (2017), assuming a scatter of 0.15 dex in $M_\star$ at a given M_{vir}. Abundance matching also provides insight on how the galaxy size is related to the host halo structure – Kravtsov (2013) and Somerville et al. (2018) found that galaxy size scales linearly with host halo virial radius, $r_{\mathrm{eff}} \sim 0.02 r_{\mathrm{vir}}$, insensitive to morphology. Jiang et al. (2019b) verified this relation in two different suites of cosmological hydro-simulations, finding that the proportionality constant does not reflect halo spin but strongly correlates with halo concentration, c_2. In particular,

$$r_{\mathrm{eff}} = 0.02(c_2/10)^{-0.07} r_{\mathrm{vir}}. \tag{B.22}$$

The dependence on halo concentration introduces a redshift and assembly history dependence into the galaxy size. We adopt this relation in order to initialize the satellite's stellar size, assuming a log-normal scatter with $\sigma = 0.15$ dex in r_{eff} at fixed r_{vir}, as found by Jiang et al. (2019b). Note that we track the evolution in the satellite's stellar half-mass radius without making any specific assumptions about the underlying density profile of the stars.

Following Zinger et al. (2018), we assume that the circumgalactic medium (CGM) of a galaxy is in hydrostatic equilibrium with the host halo and, to a good approximation, follows the halo profile according to

$$\rho_{\mathrm{gas}}(r) = f_{\mathrm{gas}} \rho(r), \tag{B.23}$$

where $f_{\rm gas}$ is the ratio of the total CGM gas mass to virial mass. For incoming satellites, we can write

$$f_{\rm gas} = \frac{f_{\rm bar}}{1 - f_{\rm bar}} - \frac{M_\star}{M_{\rm vir}}, \tag{B.24}$$

where the baryonic fraction, $f_{\rm bar}$, is given by Okamoto et al. (2008) as

$$f_{\rm bar}(M_{\rm vir}, z) = \frac{\Omega_{\rm b}}{\Omega_{\rm m}} \left\{ 1 + 0.587 \left[\frac{M_{\rm vir}}{M_{\rm c}(z)} \right]^{-2} \right\}^{-3/2}, \tag{B.25}$$

where $M_{\rm c}(z)$ is the mass below which galaxies are strongly affected by photoionization. We adopt $M_{\rm c}(z)$ from the numerical values given by Okamoto et al. (2008). This recipe implicitly assumes that supernovae feedback does not remove hot gas from the halo.

The prescriptions in §B.2.3.3 and §B.2.3.4 apply both to the central host and to the satellites at the moment of infall.

B.2.4 Orbit integration and dynamical friction

We follow the orbits by treating satellites as point masses. At each timestep, `SatGen` solves the equations of motion in the cylindrical frame using an order 4(5) Runge-Kutta method.[5] We solve

$$\ddot{\boldsymbol{r}} = -\nabla\Phi + \boldsymbol{a}_{\rm DF}, \tag{B.26}$$

where $\boldsymbol{r} = (R, \phi, z)$ is the position vector, Φ is the gravitational potential, and $\boldsymbol{a}_{\rm DF}$ is the acceleration due to dynamical friction (DF), which is modeled using the Chandrasekhar (1943) formula,

$$\boldsymbol{a}_{\rm DF} = -4\pi G^2 m \sum_i \ln\Lambda_i \, \rho_i(\boldsymbol{r}) F(< V_{{\rm rel},i}) \frac{\boldsymbol{V}_{{\rm rel},i}}{V^3_{{\rm rel},i}}. \tag{B.27}$$

Here the summation is over all of the components of the host system (e.g., i =halo, disc, and bulge, following Taylor & Babul 2001 and Peñarrubia et al. 2010), m is the instantaneous satellite mass, $\ln\Lambda_i$ is the Coulomb logarithm, $\boldsymbol{V}_{{\rm rel},i}$ is the relative velocity of the satellite with respect to the streaming motion of the particles of component i, and $F(< V_{{\rm rel},i})$ is the fraction of local host particles contributing to DF. For simplicity, we assume that the velocity distributions of all of the host components are Maxwellian and isotropic such that

$$F(< V_{{\rm rel},i}) = {\rm erf}(X_i) - \frac{2X_i}{\sqrt{\pi}} e^{-X_i^2}, \tag{B.28}$$

5. We use the 'dopri5' integrator as implemented in `scipy.integrate.ode`.

where $X_i \equiv V_{\mathrm{rel},i}/(\sqrt{2}\sigma_i)$, with $\sigma_i(\boldsymbol{r})$ the one-dimensional velocity dispersion of component i. [6]

The Coulomb logarithm and the relative velocity depend on the host component of interest. For spherical components such as the halo or bulge, we adopt $\ln\Lambda_i = \xi\ln(M_i/m)$, where the factor $\ln(M_i/m)$ is a widely used form for the Coulomb logarithm (e.g., Gan et al., 2010), with M_i and m the host mass and satellite mass, respectively, and ξ a fudge factor that accounts for the weakening of orbital decay when the density profile is cored (e.g., Read et al., 2006b). Orbital decay becomes completely stalled where the host density profile is flat, i.e., if $s = -\mathrm{d}\ln\rho/\mathrm{d}\ln r = 0$, whereas orbital decay continues where the profile is cuspy, i.e., if $s \gtrsim 1$. For simplicity, we assume $\xi = \min(s, 1)$. For discs, we use $\ln\Lambda = 0.5$, following Peñarrubia et al. (2010).

For spherical components, we use the orbital velocity $\boldsymbol{V}$ for $\boldsymbol{V}_{\mathrm{rel},i}$; i.e., we ignore the net spin of a halo or a bulge. Discs, however, have net rotation, so we use $\boldsymbol{V}_{\mathrm{rel,d}} = \boldsymbol{V} - \overline{V}_\phi\hat{\boldsymbol{\phi}}$, where the mean rotation $\overline{V}_\phi$ is given by equation (B.14).

We caution that our DF treatment is only approximate, and, as with any other attempt of modeling subhalo orbit with the Chandrasekhar (1943) formula, it carries a few conceptual inaccuracies. For instance, the Chandrasekhar (1943) formula assumed point masses moving in medium of uniform density, whereas a subhalo has an extended mass distribution and the host density along its orbit is not constant. The aforementioned choices of the Coulomb logarithm are therefore empirical corrections when extending the formula to applications beyond its assumptions. More fundamentally, Chandrasekhar (1943) considers DF to be a local effect due to the trailing gravitational wake, while DF is actually a global effect due to a response density that can operate at long distances (e.g., Weinberg, 1989). However, we have verified that the impact on satellite statistics due to this approximation is rather limited. Notably, for the experiments in Section B.4, we found that setting the disc DF term to zero only yields a $\sim$ 1% increase in the number of surviving satellites, and changing the whole $\boldsymbol{a}_{\mathrm{DF}}$ by a factor of two results in only a $\sim$10% change in the abundance of satellites.

B.2.5 Tidal stripping and ram pressure stripping

Satellites lose DM mass and stellar mass to tides, and they lose gaseous mass to ram pressure when their orbits bring them close enough to the centre of the host system.

6. In principle, for a composite potential in Jeans equilibrium and with isotropic velocity distribution, the "one-dimensional velocity dispersion of component i" (σ_i) is not well-defined, because the velocity dispersion should be calculated as a quantity for the whole system using the Jeans equation, which gives (e.g., for spherical systems): $\sigma^2(r) = G/[\sum_i\rho_i(r)]\int_\infty^r\sum_i\rho_i(r')[\sum_i M_i(r')/r'^2]\mathrm{d}r' \gtrsim \sigma_i^2(r)$. However, in practice, we find that using the σ_i of each component as if they were in equilibrium separately in isolation yields little difference in terms of the rate of orbital decay compared to using the overall $\sigma(r)$. This is mainly because $V_{\mathrm{rel},i}$ is usually larger than $\sigma(r)$, so $F(<V_{\mathrm{rel},i})$ is often not far from its maximum value of unity. Additionally, satellite mass loss and the choice of $\ln\Lambda$ both have larger impacts on DF than the detailed choice of σ. Therefore, we opt to use the σ_i of individual components, following Taylor & Babul (2001).

We estimate the instantaneous tidal radius of the satellite, $l_{\rm t}$, at each point along its orbit by solving

$$l_{\rm t} = r \left[\frac{m(l_{\rm t})/M(r)}{2 - \frac{{\rm d}\ln M(r)}{{\rm d}\ln r} + \frac{V_{\rm t}^2(\boldsymbol{r})}{V_{\rm circ}^2(r)}} \right]^{1/3} \tag{B.29}$$

(e.g., King, 1962; Taylor & Babul, 2001; Zentner & Bullock, 2003), where $m(l)$ and $M(r)$ are the enclosed mass profiles of the satellite and host, respectively, and $V_{\rm t}(\boldsymbol{r}) = |\hat{\boldsymbol{r}} \times \boldsymbol{V}|$ is the instantaneous tangential speed. The first two terms in the denominator represent the gravitational tidal force – obviously, tidal stripping depends on the local mass profile of the host (see Dekel et al. 2003 for a thorough discussion). The third term represents the differential centrifugal force across the satellite due to its orbital motion about the halo centre.

Although the tidal radius is widely used to model tidal stripping, it is an ill-defined concept for several reasons [e.g.,][](van den Bosch et al., 2018). For example, the Hill surface is not spherical or infinitesimally thin (Read et al., 2006a; Tollet et al., 2017). Because of this, we express the instantaneous mass loss rate as

$$\dot{m} = -\mathcal{A}\frac{m(>l_{\rm t})}{t_{\rm dyn}(r)}, \tag{B.30}$$

where we have introduced a fudge parameter $\mathcal{A}$ as the stripping efficiency to incapsulate uncertainties in the definition of the tidal radius. As such, the timescale on which stripping occurs is the local dynamical time $t_{\rm dyn}(r) = \sqrt{3\pi/16G\overline{\rho}(r)}$ divided by $\mathcal{A}$ (with $\overline{\rho}(r)$ the average density of the host system within radius r, including the baryonic components). We calibrate the mass loss rate model using high-resolution idealized simulations and find $\alpha \approx 0.55$ (Green et al., 2021a). [7] The mass evolution over a timestep Δt is then given by

$$m(t + \Delta t) = m(t) + \dot{m}\Delta t. \tag{B.31}$$

Similarly, if a higher-order satellite (see Fig. B.1 for definition) stays outside the tidal radius of the hosting satellite for more than a time of $t_{\rm dyn}(l)/\mathcal{A}$, where $t_{\rm dyn}(l)$ is the local dynamical time of the hosting satellite, it is released to the lower-order host, picking up a new orbital velocity that is the superposition of its velocity with respect to the previous hosting satellite and the velocity of the hosting satellite with respect to the lower-order host.

Analogous to how the tidal radius is defined, a ram pressure radius ($l_{\rm RP}$) can be defined as the satellite-

7. In several previous studies (e.g., Zentner & Bullock, 2003; Zentner et al., 2005a; Pullen et al., 2014; van den Bosch et al., 2018), the stripping time is assumed to be the instantaneous orbital time divided by a fudge factor, i.e., $(2\pi r/V_{\rm t})/A$, with $A = 1 - 6$ across the studies. Our choice of $\mathcal{A} = 0.55$ corresponds roughly to $A{\sim}1.65$ for a typical cosmological orbit, bracketed by literature values but on the inefficient-stripping end.

centric distance where the self-gravitational restoring force per unit area balances the ram pressure exerted by the gaseous host halo. We compute $l_{\rm RP}$ at each point along the orbit by solving

$$\kappa \frac{G m(l_{\rm RP}) \rho_{\rm gas}(l_{\rm RP})}{l_{\rm RP}} = \rho_{\rm gas}(\boldsymbol{r}) V(\boldsymbol{r})^2, \tag{B.32}$$

where κ is a factor of order unity (Zinger et al., 2018, $\kappa = 0.5 - 2$, depending on assumptions made in calculating the gravitational restoring force), and we take for simplicity $\kappa = 1$. The mass loss rate of the gaseous halo is given by

$$\dot{m}_{\rm gas} = -\frac{m_{\rm gas}[> \min(l_{\rm t}, l_{\rm RP})]}{2 t_{\rm dyn}(r)}. \tag{B.33}$$

In practice, $\min(l_{\rm t}, l_{\rm RP}) = l_{\rm RP}$ in most cases, i.e., ram pressure stripping is usually more efficient than tidal stripping for gas.

B.2.6 Evolution of satellite structure

Satellites react to two competing tidal effects: tidal stripping, which takes mass away and makes satellite smaller, and tidal heating, which injects orbital kinetic energy into the satellite, causing it to expand. While tidal stripping can be analytically estimated (Section B.2.5), the effect of heating, or the net structural response to both tidal effects, is not easily captured by analytical arguments. Several studies have resorted to using idealized simulations to tabulate satellite structural evolution due to the tidal field as a function of the mass that has been lost (Hayashi et al., 2003; Peñarrubia et al., 2008; Peñarrubia et al., 2010; Errani et al., 2015, 2018; Green & van den Bosch, 2019).[8] Notably, Hayashi et al. (2003) and Peñarrubia et al. (2008); Peñarrubia et al. (2010) found that subhalo density profiles depend solely on the density profile at infall and the total amount of mass lost thereafter. In particular, they describe the evolution of the maximum circular velocity ($v_{\rm max}$) and the radius at which the circular velocity reaches the maximum ($l_{\rm max}$) using a generic function,

$$g(x) = \left(\frac{2}{1+x}\right)^{\mu} x^{\eta}, \tag{B.34}$$

where $g(x) = v_{\rm max}(t)/v_{\rm max}(0)$ or $l_{\rm max}(t)/l_{\rm max}(0)$, x is the bound mass fraction ($m(t)/m(0)$), and μ and η are the best-fit parameters calibrated against idealized simulations. Peñarrubia et al. (2010) found that μ and η depend on the initial inner logarithmic density slope of the satellite, $s_{0.01}$ (see Appendix B.3.2 for their values). These relations, also known as *tidal-evolution tracks*, are scale-free, independent of the orbital parameters, and only marginally sensitive to the initial concentration of the subhaloes (Green & van den

8. But see also Du et al. (in prep), which studies the tidal heating of subhaloes using idealized N-body simulations and derives analytical formulae that accurately approximate the effects of tidal heating on subhalo density profiles.

Bosch, 2019), which we ignore here.

Errani et al. (2018) extended tidal tracks to describe the evolution of the stellar mass ($m_\star$) and half-stellar-mass radius ($l_{\rm eff}$). In particular, they found that

$$\tilde{g}(x) = \left(\frac{1+x_{\rm s}}{x+x_{\rm s}}\right)^{\mu} x^{\eta}, \tag{B.35}$$

where $\tilde{g}(x) = m_\star(t)/m_\star(0)$ or $l_{\rm eff}(t)/l_{\rm eff}(0)$ and $x = m_{\rm max}(t)/m_{\rm max}(0)$, with $m_{\rm max}$ the subhalo mass within the maximum-circular-velocity radius, $m(l_{\rm max})$. Here, the parameters, μ, η, and $x_{\rm s}$, depend not only on the initial density slope, $s_{0.01}(0)$, but also on how compact the stellar component initially is with respect to the hosting subhalo, measured by $l_{\rm eff}(0)/l_{\rm max}(0)$. Note that by using these tidal tracks, we do not assume density profiles for stellar mass or explicitly model tidal stripping of stars; instead, we updated the evolved stellar mass and half-mass radius assuming that they are coupled to the evolution of the subhaloes through $m(l_{\rm max})$. We list the parameter values in Appendix B.3.2, but summarize the tidal tracks qualitatively here as follows: satellite size generally increases with subhalo mass loss, which manifests due to tidal heating and the re-virialization response to tidal stripping and heating; only cuspy satellites ($\alpha \gtrsim 1$) can become more compact, and the size decrease occurs only after significant subhalo mass loss.

With the tidal tracks described by equations (B.34)-(B.35), the formula for tidal stripping, equations (B.29)-(B.31), and the initial profile as set up in Section B.2.3, we can completely specify the *evolved* subhalo profile, the stellar mass, and the stellar size at each timestep along the orbit. For this, a conversion between $v_{\rm max}$ and $l_{\rm max}$ and the parameters that are directly used to define a subhalo density profile, e.g., the concentration c and overdensity Δ, is needed. We provide details on such a conversion in Appendix B.3.2. For the gas distribution, we assume that the remaining gas follows the evolved subhalo profile as in equation (B.23), with $f_{\rm gas} = m_{\rm gas}(t)/m(t)$.

B.2.7 Improvements compared to previous models

`SatGen` combines the wisdom of earlier models and improves in important ways. Most previous models have focused on DM subhaloes (Taylor & Babul, 2001; Zentner & Bullock, 2003; Zentner et al., 2005a; Gan et al., 2010; Peñarrubia et al., 2010; Jiang & van den Bosch, 2016), whereas `SatGen` takes baryonic properties into consideration. A couple of models have included certain details of baryonic processes (Carleton et al., 2019; Nadler et al., 2019), but `SatGen` is more thorough.

For example, the model by Nadler et al. (2019) considers the stellar component. It initializes the satellite stellar size in the same way as `SatGen` , but for the size evolution it only considers size decrease due to tidal stripping and neglects expansion due to tidal heating, which is a process that is essential for producing

UDGs in dense environments (Carleton et al., 2019; Jiang et al., 2019a). Also, tidal stripping in this model is treated in an orbit-averaged sense, as in van den Bosch et al. (2005b) and Jiang & van den Bosch (2016). This treatment washes out detailed mass and structural evolution along the orbits.

The model by Carleton et al. (2019) uses the same tidal tracks as used in `SatGen`; however, it applies abrupt tidal truncation to satellites at orbital pericentres such that pericentres are the only locations where the satellites lose mass. This is not accurate for circular orbits or any orbits with $\epsilon \gtrsim 0.5$. In addition, the Carleton et al. (2019) model relies on cosmological N-body simulations for merger trees, orbits, and initial conditions. In contrast, `SatGen` can generate larger samples using the EPS formalism, which is useful for studying the halo-to-halo variance of satellite properties, and can follow the orbits self-consistently.

B.2.8 Illustration and workflow

We present an idealized example of a massive satellite orbiting a MW-sized halo in Section B.3.3 in order to provide an intuitive illustration (Fig. B.3) of the orbit integration and satellite evolution prescriptions described in Section B.2.2-Section B.2.6.

When using `SatGen` for a cosmological setup, we summarize the workflow as follows:

1. Starting with a target halo of a given mass and redshift, draw halo merger trees according to Section B.2.1.
2. Initialize host and satellite properties according to Section B.2.3, using density profiles introduced in Section B.2.2 and Section B.3.1, and considering halo response models that are characteristic of certain cosmological hydro-simulations.
3. Evolve the satellites: integrate the orbit according to Section B.2.4 and update the masses and profiles of the satellites and the host for every timestep of $\Delta t = 0.1 t_{\rm dyn}(z)$, according to Section B.2.6.

This procedure is somewhat similar to that of zoom-in simulations, in the sense that both `SatGen` and zoom-in simulations start with a target halo and then trace the progenitors back in time, finally evolving forward in time to refine the small-scale structures.

B.3 Supplementary model details

B.3.1 Analytics of profiles

Here, we provide the analytical expressions for the profiles of density (ρ), enclosed mass (M), gravitational potential (Φ), the R-component and z-component of gravitational acceleration in the cylindrical coordinate system (f_R, f_z), and the one-dimensional velocity dispersion for an isotropic velocity distribution (σ), as

well as a few convenient relations among the parameters, for each of the potential well classes supported in `SatGen`.

B.3.1.1 NFW

We specify an NFW profile using the virial mass, $M_{\rm vir}$, the concentration parameter, c_2 (or the corresponding scale radius $r_{\rm s} = r_{\rm vir}/c_2$), and the average spherical overdensity, Δ.

$$\rho(r) = \frac{\rho_0}{x\,(1+x)^2}, \quad \text{where} \quad x = \frac{r}{r_{\rm s}} \quad \text{and} \quad \rho_0 = \frac{c_2^3}{3f(c_2)}\Delta\rho_{\rm crit}, \tag{B.36}$$

with $f(x) = \ln(1+x) - x/(1+x)$.

$$M(r) = M_{\rm vir}\frac{f(x)}{f(c_2)}. \tag{B.37}$$

$$\Phi(r) = \Phi_0\frac{\ln(1+x)}{x}, \quad \text{where} \quad \Phi_0 = -4\pi G\rho_0 r_{\rm s}^2. \tag{B.38}$$

$$f_R = -\frac{\partial\Phi}{\partial R} = \Phi_0\frac{f(x)}{x}\frac{R}{r^2} \quad \text{and} \quad f_z = -\frac{\partial\Phi}{\partial z} = \Phi_0\frac{f(x)}{x}\frac{z}{r^2}, \tag{B.39}$$

where $r = \sqrt{R^2+z^2}$.

$$\begin{aligned}\sigma^2(r) &= V_{\rm vir}^2\frac{c}{f(c)}x(1+x)^2\int_x^\infty\frac{f(x')}{x'^3(1+x')^2}{\rm d}x' \\ &\approx V_{\rm max}^2\left(\frac{1.4393x^{0.354}}{1+1.1756x^{0.725}}\right)^2,\end{aligned} \tag{B.40}$$

where the second line is an approximation accurate to 1% for $x = 0.01$-100 (Zentner & Bullock, 2003, see also an analytical solution involving non-elementary functions by Łokas & Mamon 2001).

The location of the peak circular velocity, $r_{\rm max}$, is related to the scale radius, $r_{\rm s}$, by

$$r_{\rm max} \approx 2.163 r_{\rm s}, \tag{B.41}$$

where $r_{\rm s}$ is the location at which the logarithmic density slope is 2, r_2.

B.3.1.2 Dekel+

We specify a Dekel+ profile using the virial mass, $M_{\rm vir}$, a concentration parameter, c (or the corresponding scale radius $r_{\rm s} = r_{\rm vir}/c$), the innermost logarithmic density slope, $\alpha \equiv -{\rm d}\ln\rho/\ln r|_{r\to 0}$, and the average

spherical overdensity, Δ.

$$\begin{aligned} \rho(r) &= \frac{\rho_0}{x^\alpha (1+x^{1/2})^{2(3.5-\alpha)}}, \\ \text{where } x &= \frac{r}{r_\mathrm{s}} \text{ and } \rho_0 = \frac{c^3(3-\alpha)}{3f(c,\alpha)}\Delta\rho_\mathrm{crit}, \end{aligned} \tag{B.42}$$

with $f(x,\alpha) = \chi^{2(3-\alpha)}$ and $\chi = x^{1/2}/(1+x^{1/2})$.

$$M(r) = M_\mathrm{vir}\frac{f(x,\alpha)}{f(c,\alpha)}. \tag{B.43}$$

$$\Phi(r) = -V_\mathrm{vir}^2 \frac{2c}{f(c,\alpha)}\left[\frac{1-\chi^{2(2-\alpha)}}{2(2-\alpha)} - \frac{1-\chi^{2(2-\alpha)+1}}{2(2-\alpha)+1}\right]. \tag{B.44}$$

$$\begin{aligned} f_R(R,z) &= (2-\alpha)[2(2-\alpha)+1]\Phi_0 \frac{f(x,\alpha)}{x}\frac{R}{r^2} \quad \text{and} \\ f_z(R,z) &= (2-\alpha)[2(2-\alpha)+1]\Phi_0 \frac{f(x,\alpha)}{x}\frac{z}{r^2}, \\ \text{where } \Phi_0 &= -\frac{4\pi G\rho_0 r_\mathrm{s}^2}{(3-\alpha)(2-\alpha)[2(2-\alpha)+1]}. \end{aligned} \tag{B.45}$$

$$\begin{aligned} \sigma^2(r) &= V_\mathrm{vir}^2 \frac{c}{f(c,\alpha)}\frac{x^{3.5}}{\chi^{2(3.5-\alpha)}}\int_x^\infty \frac{\chi(x')^{4(3-\alpha)+1}}{x'^{5.5}}\mathrm{d}x' \\ &= 2V_\mathrm{vir}^2 \frac{c}{f(c,\alpha)}\frac{x^{3.5}}{\chi^{2(3.5-\alpha)}}\sum_{i=0}^{8}\frac{(-1)^i 8!}{i!(8-i)!}\frac{1-\chi^{4(1-\alpha)+i}}{4(1-\alpha)+i}. \end{aligned} \tag{B.46}$$

We refer interested readers to Freundlich et al. (2020b) for the analytical expressions of the Dekel+ profile for gravitational lensing-related quantities, including the surface density, deflection angle, shear, and magnification.

Unlike NFW, for which $r_\mathrm{s} = r_2$, the Dekel+ scale radius is related to r_2 by

$$r_2 = r_s\left(\frac{2-\alpha}{1.5}\right)^2, \tag{B.47}$$

such that the relation between the Dekel+ concentration (c) and the conventional concentration (c_2) is

$$c_2 = \frac{r_\mathrm{vir}}{r_2} = \left(\frac{1.5}{2-\alpha}\right)^2 c. \tag{B.48}$$

The location of peak circular velocity, r_max, is related to r_2 by

$$r_\mathrm{max} = 2.25 r_2 = (2-\alpha)^2 r_\mathrm{s}. \tag{B.49}$$

The profile of the logarithmic density slope is

$$s(r) = -\frac{\mathrm{d}\ln\rho}{\mathrm{d}\ln r} = \frac{\alpha + 3.5\sqrt{x}}{1+\sqrt{x}}. \tag{B.50}$$

The slope at $0.01r_{\rm vir}$ is

$$s_{0.01} \equiv s(0.01r_{\rm vir}) = \frac{\alpha + 0.35\sqrt{c}}{1+0.1\sqrt{c}}. \tag{B.51}$$

For $s_{0.01}$ values that are commonly seen in simulations and observations $(0-2)$ and for a typical concentration (e.g., $c = 10$), we have $\alpha \in (-1.11, 1.53)$. That is, α can be negative for realistic profiles, and thus $s_{0.01}$ is a more physically meaningful quantity than α when it comes to comparing the cuspiness of density profiles.

B.3.1.3 Einasto

We define an Einasto profile using the virial mass, $M_{\rm vir}$, the concentration parameter, c (or the corresponding scale radius $r_{\rm s} = r_{\rm vir}/c$), the shape index, n, and the average spherical overdensity, Δ.

$$\begin{aligned} &\rho(r) = \rho_0 e^{-x(r)}, \\ &\text{where } \; x = 2n\left(\frac{r}{r_{\rm s}}\right)^{\frac{1}{n}} \; \text{ and } \; \rho_0 = \frac{M_{\rm vir}}{4\pi h^3 n\gamma[3n, x(r_{\rm vir})]}, \end{aligned} \tag{B.52}$$

with $h = r_{\rm s}/(2n)^n$ and $\gamma(a,x)$ is the non-normalized lower incomplete gamma function. Here, we have adopted the notations in Retana-Montenegro et al. (2012) for compact expressions.

$$M(r) = M_{\rm tot}\tilde{\gamma}(3n,x), \;\; \text{with} \;\; M_{\rm tot} = 4\pi\rho_0 h^3 n\Gamma(3n), \tag{B.53}$$

where $\Gamma(a)$ and $\tilde{\gamma}(a,x) = \gamma(a,x)/\Gamma(a)$ are the Gamma function and the normalized lower incomplete gamma function, respectively.

$$\Phi(r) = -\frac{GM_{\rm tot}}{h}\left[\frac{\tilde{\gamma}(3n,x)}{x^n} + \frac{\Gamma(2n,x)}{\Gamma(3n)}\right], \tag{B.54}$$

where $\Gamma(a,x)$ is the non-normalized upper incomplete gamma function.

$$\begin{aligned} f_R(R,z) &= -GM_{\rm tot}\tilde{\gamma}(3n,x)\frac{R}{r^3} \quad \text{and} \\ f_z(R,z) &= -GM_{\rm tot}\tilde{\gamma}(3n,x)\frac{z}{r^3}. \end{aligned} \tag{B.55}$$

$$\sigma^2(r) = \frac{GM_{\rm tot}}{h} ne^x \int_x^\infty \frac{\tilde{\gamma}(3n,x')}{e^{x'}x'^{(n+1)}}\mathrm{d}x'. \tag{B.56}$$

Like the NFW profile, the Einasto scale radius, r_s, is the same as r_2, where the logarithmic density slope is 2. The radius of peak circular velocity is related to r_s by

$$r_{max} \approx 1.715\alpha^{-0.00183}(\alpha + 0.0817)^{-0.179488} r_s \tag{B.57}$$

(Garrison-Kimmel et al., 2014b). The profile of the logarithmic density slope is

$$s(r) = -\frac{\mathrm{d}\ln\rho}{\mathrm{d}\ln r} = \frac{x(r)}{n}, \tag{B.58}$$

so

$$s_{0.01} = 2(0.01c)^{\frac{1}{n}}. \tag{B.59}$$

B.3.1.4 MN

We define a MN profile using the disc mass, M_d, a scale radius, a, and a scale height, b.

$$\rho(R,z) = \frac{M_d b^2}{4\pi} \frac{aR^2 + (a+3\zeta)(a+\zeta)^2}{\zeta^3[R^2 + (a+\zeta)^2]^{5/2}}, \tag{B.60}$$

where $\zeta = \sqrt{z^2 + b^2}$.

$$M(r) = \frac{M_d r^3}{[r^2 + (a+b)^2]^{1.5}}, \quad \text{where} \quad r = \sqrt{R^2 + z^2}. \tag{B.61}$$

$$\Phi(R,z) = -\frac{GM_d}{\sqrt{R^2 + (a+\zeta)^2}}. \tag{B.62}$$

$$\begin{aligned} f_R(R,z) &= -\frac{GM_d}{[R^2 + (a+\zeta)^2]^{1.5}} R \quad \text{and} \\ f_z(R,z) &= -\frac{GM_d}{[R^2 + (a+\zeta)^2]^{1.5}} \frac{a+\zeta}{\zeta} z. \end{aligned} \tag{B.63}$$

$$\sigma^2(R,z) = \frac{GM_d^2 b^2}{8\pi\rho(R,z)} \frac{(a+\zeta)^2}{\zeta^2[R^2 + (a+\zeta)^2]^3}. \tag{B.64}$$

The relation between half-mass radius, r_{eff}, and the scale lengths, (a, b), is

$$a = \frac{0.766421}{1 + b/a} r_{eff}. \tag{B.65}$$

B.3.2 Structure of evolved satellites

B.3.2.1 Tidal evolution tracks

We use the tidal evolution tracks of Peñarrubia et al. (2010) for determining the profiles of evolved subhaloes and those of Errani et al. (2018) for updating the stellar masses and half-stellar-mass radii. These tidal tracks can be expressed with the universal functional form of

$$g(x) = \left(\frac{1+x_\mathrm{s}}{x+x_\mathrm{s}}\right)^{\mu} x^{\eta}, \tag{B.66}$$

where, for the DM subhalo, g represents $v_\mathrm{max}(t)/v_\mathrm{max}(0)$ or $l_\mathrm{max}(t)/l_\mathrm{max}(0)$, and x stands for the bound mass fraction $m(t)/m(0)$. For the stellar component, g represents $m_\star(t)/m_\star(0)$ or $l_\mathrm{eff}(t)/l_\mathrm{eff}(0)$, and x stands for $m_\mathrm{max}(t)/m_\mathrm{max}(0)$, with $m_\mathrm{max} = m(l_\mathrm{max})$. The parameters μ and η depend on the initial logarithmic density slope, $s_{0.01}$ ($\equiv -\mathrm{d}\ln\rho/\mathrm{d}\ln r|_{r=0.01r_\mathrm{vir}}$), and x_s depends on the initial stellar size with respect to the initial radius of peak circular velocity of the hosting subhalo, $l_\mathrm{eff}(0)/l_\mathrm{max}(0)$. Peñarrubia et al. (2010) and Errani et al. (2018) obtained best-fit parameters for different initial structures ($s_{0.01} = 0, 0.5, 1, 1.5$ and $l_\mathrm{eff}(0)/l_\mathrm{max}(0) = 0.05, 0.1$) by calibrating the model against idealized N-body simulations, which we summarize here in Table B.1. For the initial structures not listed in the table but within the range of the tabulated initial structures, we use cubic spline interpolation to get the parameters. For the initial structures beyond the tabulated range, we do not extrapolate, but use the nearest neighbours in the table.

Fig. B.2 illustrates these tidal tracks. Note that stellar mass loss is marginal when the subhalo mass within l_max decreases by $\lesssim 90\%$, especially when the initial stellar mass distribution is compact (e.g., when $l_\mathrm{eff}(0)/l_\mathrm{max}(0) = 0.05$). Also note that, generally, satellite size increases with subhalo mass loss, which manifests due to tidal heating and re-virialization in response to tidal stripping and heating. Only cuspy satellites ($\alpha \gtrsim 1$) become more compact in stellar size, and the size decrease occurs only after significant subhalo mass loss, when $m_\mathrm{max}(l_\mathrm{max})$ decreases by $\gtrsim 99\%$. This is, however, a viable channel for making compact bright dwarfs ($m_\star \sim 10^{7-9}$ and $l_\mathrm{eff} \lesssim 1$ kpc) from massive cuspy galaxies.

B.3.2.2 Evolved subhalo profiles

The parameters that we use to define the subhalo profiles – e.g., for the Dekel+ profile – c, α, and Δ, are not directly provided by the tidal tracks. We need to translate $(v_\mathrm{max}, l_\mathrm{max})$ to (c, α, Δ) in order to update the profiles of evolved subhaloes. Note that the evolved subhaloes have higher overdensities (Δ) compared to distinct haloes, which all have $\Delta = 200$.

Since the number of parameters (c, α, Δ) exceeds that of the constraints $(v_\mathrm{max}, l_\mathrm{max})$, we need an ad-

initial structure \ quantity		$g(x)=\frac{v_{\rm max}(t)}{v_{\rm max}(0)}, x=\frac{m(t)}{m(0)}$			$g(x)=\frac{l_{\rm max}(t)}{l_{\rm max}(0)}, x=\frac{m(t)}{m(0)}$			$g(x)=\frac{l_{\rm eff}(t)}{l_{\rm eff}(0)}, x=\frac{m_{\rm max}(t)}{m_{\rm max}(0)}$			$g(x)=\frac{m_\star(t)}{m_\star(0)}, x=\frac{m_{\rm max}(t)}{m_{\rm max}(0)}$		
$s_{0.01}$	$\frac{l_{\rm eff}(0)}{l_{\rm max}(0)}$	μ	η	$\log x_{\rm s}$	μ	η	$\log x_{\rm s}$	μ	η	$\log x_{\rm s}$	μ	η	$\log x_{\rm s}$
1.5	0.05	0.4	0.24	0	0	0.48	0	(0.59)	(0.59)	(-2.4)	(1.39)	(1.39)	(-2.4)
	0.1							(0.75)	(0.71)	(-2.0)	(1.68)	(1.68)	(-2.0)
1	0.05	0.4	0.3	0	-0.3	0.4	0	0.47	0.41	-2.64	1.87	1.87	-2.64
	0.1							0.5	0.42	-2.08	1.8	1.8	-2.08
0.5	0.05	0.4	0.35	0	-0.4	0.27	0	(0.19)	(0.07)	(-2.9)	(2.35)	(2.35)	(-2.9)
	0.1							(0.21)	(0.09)	(-2.2)	(1.93)	(1.93)	(-2.2)
0	0.05	0.4	0.37	0	-1.3	0.05	0	-0.15	-0.35	-3.12	2.83	2.83	-3.12
	0.1							-0.15	-0.33	-2.33	2.05	2.05	-2.33

Table B.1: Tidal-evolution tracks of the functional form $g(x) = [(1+x_{\rm s})/(x+x_{\rm s})]^{\mu} x^{\eta}$, compiled from Peñarrubia et al. (2010) and Errani et al. (2018) – for subhaloes, g represents $v_{\rm max}(t)/v_{\rm max}(0)$ or $l_{\rm max}(t)/l_{\rm max}(0)$ and x stands for the bound mass fraction, $m(t)/m(0)$; for stellar components, g represents $m_\star(t)/m_\star(0)$ or $l_{\rm eff}(t)/l_{\rm eff}(0)$, and x stands for $m_{\rm max}(t)/m_{\rm max}(0)$, where $m_{\rm max} = m(l_{\rm max})$. The parameter values in brackets are from linear interpolation/extrapolation.

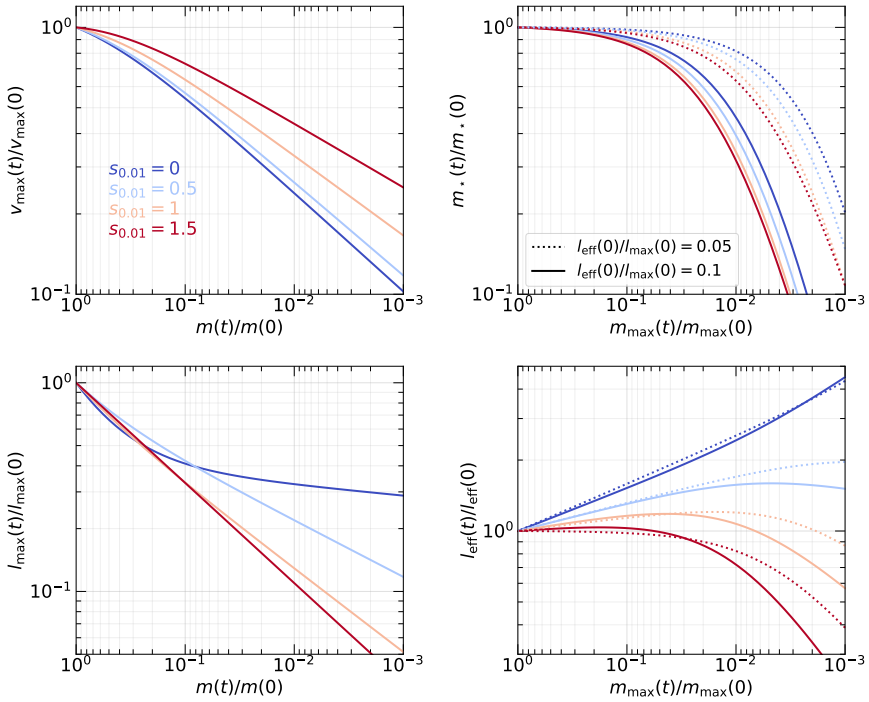

Figure B.2: Tidal evolution tracks, compiled from Peñarrubia et al. (2010) and Errani et al. (2018) – instantaneous subhalo v_{max} and l_{max} in units of their initial values, both as functions of the instantaneous bound mass fraction, $m(t)/m(0)$ (*left*); instantaneous stellar mass, m, and half-stellar-mass radius, l_{eff}, in units of their initial values, both as functions of the instantaneous ratio between the subhalo mass within l_{max} (i.e., $m_{max} \equiv m(l_{max})$) and the initial value of m_{max}. The tracks depend on the initial inner density slope ($s_{0.01}$), and for the stellar component, also depend on the initial compactness of the stellar distribution (as parameterized by $l_{eff}(0)/l_{max}(0)$).

ditional assumption. We follow Peñarrubia et al. (2010) to assume that the innermost slope α is constant. One can analytically show that the innermost part of a subhalo is adiabatically shielded against tidal shocks (Gnedin et al., 1999a). In addition, several numerical studies have shown that the logarithmic density slope at $l \longrightarrow 0$ barely changes even if the subhalo is stripped down to 0.1% of its initial mass (Peñarrubia et al., 2010; van den Bosch et al., 2018; van den Bosch & Ogiya, 2018). Under this assumption, we can express c and Δ in terms of $v_{\rm max}$ and $l_{\rm max}$. We use two relations, $\mathrm{d}v^2_{\rm circ}/\mathrm{d}l|_{l_{\rm max}} = 0$ and $v^2_{\rm max} = v^2_{\rm circ}(l_{\rm max})$, which give

$$c = (2-\alpha)^2 \frac{l_{\rm vir}}{l_{\rm max}} \tag{B.67}$$

and

$$v^2_{\rm max} = \frac{G m_{\rm vir}}{l_{\rm max}} \frac{f[(2-\alpha)^2, \alpha]}{f(c,\alpha)}, \tag{B.68}$$

where $f(x,\alpha) = \chi^{2(3-\alpha)}$ and $\chi = x^{1/2}/(1+x^{1/2})$. Combining these two relations, we can express the evolved virial mass ($l_{\rm vir}$) and thus the evolved overdensity (Δ) in terms of $m_{\rm vir}$, α, $v_{\rm max}$, and $l_{\rm max}$ as

$$\Delta = \frac{3 m_{\rm vir}}{4\pi l^3_{\rm vir} \rho_{\rm crit}(z)}, \tag{B.69}$$

and

$$l_{\rm vir} = \frac{l_{\rm max}}{(2-\alpha)^2} \frac{\chi^2_c}{(1-\chi_c)^2}, \quad \text{with} \quad \chi_c = \left(\frac{G m_{\rm vir}}{l_{\rm max} v^2_{\rm max}}\right)^{\frac{1}{2(3-\alpha)}} \left(\frac{2-\alpha}{3-\alpha}\right). \tag{B.70}$$

Using equations (B.67), (B.69), and (B.70), we can update an evolved Dekel+ subhalo according to the mass $m_{\rm vir}(t)$ from the tidal stripping recipe in Section B.2.5 and the evolved structure, $l_{\rm max}$ and $v_{\rm max}$, from the tidal tracks.

One can derive equivalent expressions for the Einasto profile, linking the Einasto concentration, c, the shape index, n, and the overdensity, Δ, to $v_{\rm max}$, $l_{\rm max}$, and an inner slope, $s(10^{-3} l_{\rm vir}) = 2(10^{-3} c)^{1/n}$, which is assumed to be constant. We omit the derivations here.

B.3.3 Illustration: evolution of one satellite in a constant potential

As an illustration of what has been described in Section B.2.2-Section B.2.6, Fig. B.3 presents the evolution of a satellite in a fixed host potential consisting of a Dekel+ halo and a MN disc. The satellite initially has a halo mass of $m_{\rm vir} = 10^{11}\,\mathrm{M}_\odot$ and is described by a Dekel+ profile with $c = 20$ and $\alpha = 0$, which corresponds to a conventional concentration of $c_2 = 11.25$ and an inner density slope of $s_{0.01} \approx 1.08$. It is also initialized with a stellar mass of $m_\star = 10^9\,\mathrm{M}_\odot$ and a half-stellar-mass radius of $l_{\rm eff} = 1.6$ kpc. The central galaxy has a halo of $M_{\rm vir} = 10^{12}\,\mathrm{M}_\odot$, $c = 10$, and $\alpha = 0.5$ (i.e., $c_2 = 10$ and $s_{0.01} = 1.22$), as well as a disc of mass

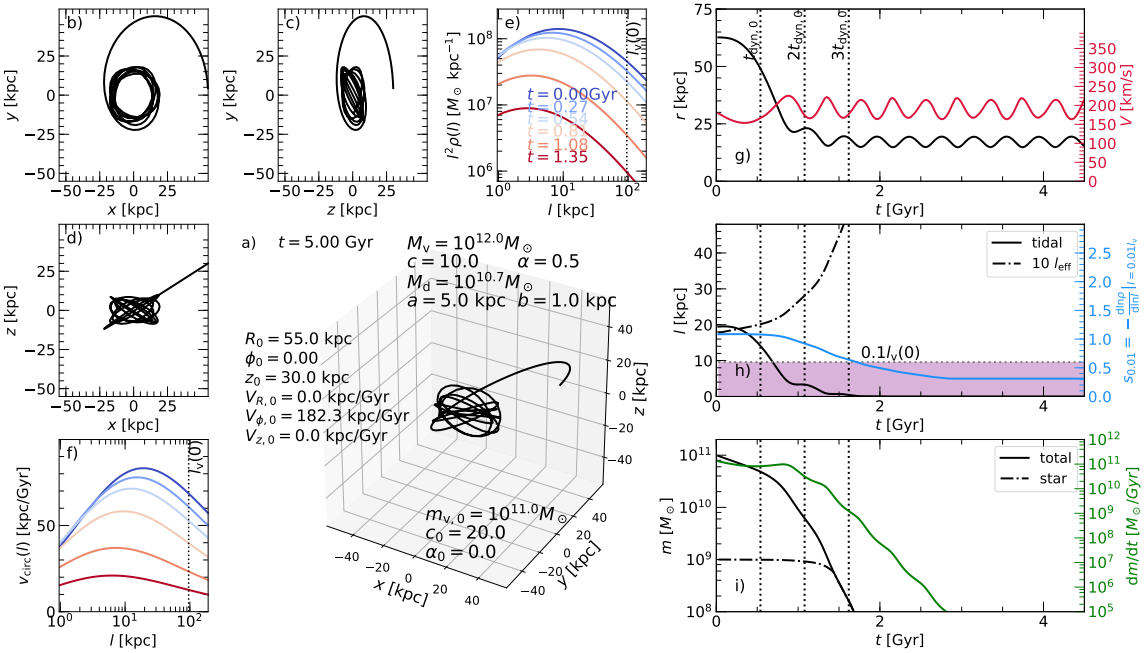

Figure B.3: Illustration of satellite evolution in `SatGen`: an idealized case where a satellite with initial halo mass of $m_{\rm vir} = 10^{11}$ $M_\odot$ described by a Dekel+ profile with $c = 20$ and $\alpha = 0$ (i.e., $c_2 = 11.25$ and $s_{0.01} \approx 1.1$) orbits around a central galaxy consisting of a halo of $M_{\rm vir} = 10^{12}$ $M_\odot$, $c = 10$, and $\alpha = 0.5$ (i.e., $c_2 = 10$ and $s_{0.01} = 1.22$) and a disc of mass $M_{\rm d} = 10^{10.7}$ $M_\odot$ with a scale size of $a = 5$ kpc and a scale height of $b = 1$ kpc. The satellite is released from $(R, z) = (55, 30)$ with a $\hat{\boldsymbol{\phi}}$-direction velocity of approximately the local circular velocity of the host potential and is evolved for 5 Gyr, during which the host potential is fixed (see the text for more details). Panels (a)-(d) show the orbit in 3D and in the$x - y$, $y - z$, and $x - z$ planes, respectively. Panels (e)-(f) show the density profile and circular velocity profile at different epochs, as indicated. The initial virial radius of the satellite is marked by the vertical dotted line. Panels (g)-(i) show the instantaneous values of a few quantities of the satellite as functions of time – (g) orbital radius and orbital velocity; (h) tidal radius, half-stellar-mass radius, and logarithmic density slope at $0.01 l_{\rm vir}(t)$ (the horizontal dotted line indicates 10% of the initial virial radius; once the tidal radius drops below this line, the stellar mass loss becomes significant); (i) subhalo mass, stellar mass, and the subhalo mass loss rate. As a massive satellite, it experiences strong dynamical friction such that its orbit decays by roughly two-thirds in radius in ∼2 initial, local dynamical times or ∼1 Gyr [Panel (e)]. It experiences tidal stripping and structural evolution along the way: notably, the maximum circular velocity decreased by roughly one third [Panel (f): the solid lines show the $v_{\rm circ}(l)$ profiles]; the half-stellar-mass radius increased by 50% [Panel (h), dash-dotted line]; the inner density slope ($s_{0.01}$) decreased from 1.1 to 0.3 [Panel (h), blue line]. Afterwards, the disc dominates the dynamics, working to drag the satellite into co-rotation.

$M_{\rm d} = 10^{10.7}\,{\rm M}_\odot$ with a scale size of $a = 5$ kpc and a scale height of $b = 1$ kpc. The satellite is released from an off-disc-plane position, $(R, z) = (55, 30)$, with an initial velocity that is approximately the local circular velocity in the $\hat{\boldsymbol{\phi}}$ direction. All of these are arbitrary choices for illustration purposes.

As can be expected, this massive satellite, with a satellite-to-central mass ratio of ∼0.1, experiences strong dynamical friction. In about two initial, local dynamical times (∼1 Gyr), its orbital radius decays from the initial ∼60 kpc to $\lesssim$ 20 kpc, where it experiences strong tidal stripping, with the instantaneous tidal radius dropping below 10% of its initial virial radius. Tidal stripping, heating, and the re-virialization of the satellite is captured by the tidal evolutionary tracks, such that after the ∼1 Gyr evolution: first, the density profile becomes shallower at $0.01 l_{\rm vir}$; second, the maximum circular velocity, $v_{\rm max}$, drops from ∼90 to ∼60 kpc/Gyr, and the $v_{\rm max}$ location, $l_{\rm max}$, decreases from 20 kpc to 8 kpc; finally, the half-stellar-mass radius increases from 1.6 kpc to 2.5 kpc.

Afterwards, the strong mass loss weakens the dynamical friction force and the influence of the disc begins to kick in: the dynamical friction force from the disc works to to drag the satellite into co-rotation, such that after traversing the disc plane several times, the satellite gradually settles into a stable orbit with a radius between 15 and 20 kpc.

B.4 Satellites of MW/M31 sized host haloes

For a proof-of-concept application, we use `SatGen` to generate satellite galaxies for MW/M31-sized host systems, studying baryonic effects on satellite statistics including subhalo abundance, spatial distribution, and internal structures. In particular, we highlight the impact of two separate baryonic effects. The first is the impact that (supernova) feedback can have on the central density profile of the (sub)haloes hosting satellites. We refer to this as the *internal* effect due to baryons. The second is the impact that the baryonic disc of the host system has on the orbital and tidal evolution of satellites. In what follows we refer to these as the *internal* and *external* baryonic effects, respectively.

B.4.1 Model setup and satellite statistics

We consider two different halo response models, which are representative of simulations of bursty star formation and strong supernovae feedback, such as NIHAO (Wang et al., 2015) and FIRE (Hopkins et al., 2014, 2018), and of simulations of non-bursty star formation and weaker feedback, such as APOSTLE (Sawala et al., 2016) and Auriga (Grand et al., 2017). We denote these two models as the NIHAO emulator and APOSTLE emulator, respectively, and tabulate the parameters of their halo response curves, as in equations (B.19)-(B.20), in Table B.2.

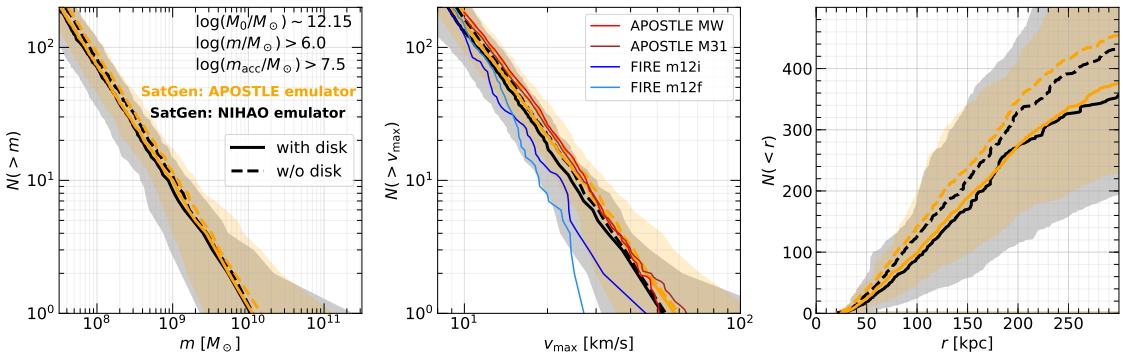

Figure B.4: Satellite statistics predicted by `SatGen` in NIHAO- and APOSTLE-emulating modes – the cumulative subhalo mass function $N(>m)$ (*left*), subhalo $v_{\rm max}$ function $N(>v_{\rm max})$ (*middle*), and radial distribution $N(<r)$ (*right*) of all of the surviving satellites in MW/M31-sized hosts (where "surviving" means $m > 10^6 \mathrm{M}_\odot$ at $z = 0$ and "MW/M31-sized" means that the present-day host halo mass is in the range $M_0 = 10^{12-12.3}\,\mathrm{M}_\odot$; see Section B.4.1 for details). Thick lines represent the median model predictions, with solid and dashed lines differentiating the cases with and without a disc potential. The colors differentiate results from the NIHAO emulator (black) and the APOSTLE emulator. Shaded bands indicate halo-to-halo variance (3-97 percentiles). The thin lines in the middle panel are APOSTLE and FIRE simulation results for the $v_{\rm max}$ function (Sawala et al., 2016; Garrison-Kimmel et al., 2017). Halo response differences result in a relatively minor effect: the NIHAO-like feedback yields $\sim$ 5% fewer satellites than the APOSTLE-like model. A baryonic disc reduces the abundance of surviving satellites within 300 (100) kpc by $\sim$ 20% (30%). Both baryonic effects are weak compared to the halo-to-halo variance.

	NIHAO emulator [a]	APOSTLE emulator
equation (B.19)	*for concentration*	
a_0	1.14	1
a_1	186	186
a_2	1	0
b_1	1.37	–
b_2	0.142	–
equation (B.20)	*for inner density slope*	
n_0	1.45	1.45
n_1	1	1
X_0	2.54×10^{-3}	2.54×10^{-3}
X_1	9.87×10^{-4}	–
ξ_0	0.21	0.21
ξ_1	2.14	0

[a] Freundlich et al. (2020b).

Table B.2: Halo response relations adopted by the two simulation emulators considered in Section B.4.

For each emulator, we randomly generate 100 merger trees for MW- and M31-sized haloes ($M_{\rm vir} = 10^{12-12.3}\,{\rm M}_\odot$ at $z = 0$), recording progenitor haloes down to $10^{7.5}\,{\rm M}_\odot$ up to $z = 20$. We initialize the satellites and hosts as described in Section B.2.3 – at this stage, the halo response relations are taken into account.[9] We then evolve the satellites, considering two cases. In one case, the host potential is just a DM halo following the Dekel+ profile, as determined by the merger tree and the initialization procedure. In the other case, the host potential consists of both the DM halo and a galactic disc. The disc mass is set to be 0.1 times the instantaneous halo mass, i.e., $M_{\rm d}(z) = 0.1 M_{\rm vir}(z)$. The disc follows a MN profile with $b/a = 1/25$. The disc size, a, is determined using the half-mass radius, $r_{\rm eff}$, as given by equation (B.22), and the relation between the MN a and $r_{\rm eff}$, as given by equation (B.65). Our discs are similar to those of Peñarrubia et al. (2010) in terms of mass and axis ratio. While approximately mimicking the cold discs of the MW or M31, these parameters are chosen mainly for illustration purposes and are not intended to reproduce the actual discs in the MW or M31 in any detail. In fact, they are on the massive side of the observationally-inferred values (e.g., Sofue, 2013).

In total we have four suites of simulations for a total of 400 MW/M31 sized haloes – we have two suites for each simulation emulator and, for each emulator, we consider the case with and without the embedded galactic disc. The merger trees and initial satellite structures of the with-disc and no-disc models are identical. This enables us to quantify the disc effect.

Fig. B.4 presents the cumulative subhalo mass functions, $N(>m)$, subhalo $v_{\rm max}$ functions, $N(>v_{\rm max})$, and satellite galactocentric-distance distributions, $N(<r)$, for all of the surviving satellites in the four suites at $z = 0$. Here, we define "surviving" as having subhalo mass larger than $10^6\,{\rm M}_\odot$ and have verified that our results are not sensitive to this arbitrary mass threshold. Lines represent the median mass, $v_{\rm max}$, or distance at fixed number N, and the shaded bands indicate the 3-97 percentiles, reflecting the halo-to-halo variance due to random assembly histories. We overplot the $v_{\rm max}$ functions from the FIRE and APOSTLE simulations, finding that the `SatGen` predictions are in reasonable agreement with the simulation results. We emphasize that this agreement is achieved without tuning any of the model parameters. We think that given the differences among the simulations, the halo-to-halo scatter, and the concern on the reliability of the simulation results due to numerical disruption (van den Bosch et al., 2018), there is no need to fine-tune the model to match the simulations in detail.

The census of bright satellites ($m_\star > 10^5\,{\rm M}_\odot$) of MW and M31 is relatively complete (e.g., Tollerud et al., 2008), so we use them as our observational benchmarks. Fig. B.5 presents the `SatGen` $v_{\rm max}$ functions and radial distributions for the massive surviving satellites with $m_\star > 10^5\,{\rm M}_\odot$ at $z = 0$, and compares them

9. For this proof-of-concept study, we opt to only follow the DM and stellar components, ignoring the gaseous components.

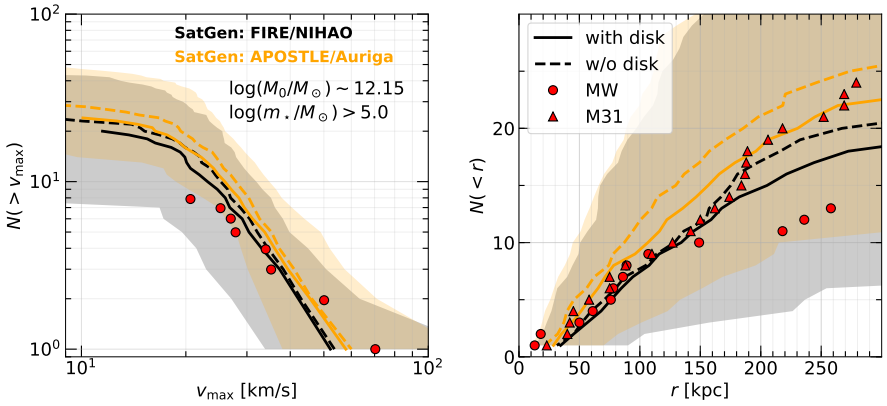

Figure B.5: Subhalo $v_{\rm max}$ functions and radial distributions of massive satellites ($m_\star > 10^5$ M$_\odot$), comparing model predictions for MW/M31-sized host haloes (lines) and observations of the actual MW/M31 (symbols). The shaded areas indicate halo-to-halo variance (3-97 percentiles, for the models with discs). The flattening of the $v_{\rm max}$ function at the low-mass compared to the middle panel of Fig. B.4 is simply due to the stellar mass cut. The APOSTLE-like feedback on average yields ~25% more massive satellites than the NIHAO-like feedback, illustrating that cuspier and denser satellites are more resistant to tidal stripping and heating. The NIHAO emulator prediction of the median radial distribution agrees well with the observations out to ~150 kpc.

with those of the McConnachie (2012) observational sample of MW/M31 satellites. We find that the model predictions agree well with those of the actual MW/M31 satellites. Notably, the median radial distribution from the NIHAO emulator agrees with the MW and M31 observations at percent-level out to ~ 150 kpc from the galactic centre, and even the observational results at the outskirts are well within the halo-to-halo variance of the model predictions.

B.4.2 Effects of different baryonic physics

In `SatGen`, the effect of different sub-grid baryonic physics is captured by the halo response relations (Section B.2.3). Among high-resolution cosmological simulations, NIHAO and FIRE feature bursty star formation histories and thus strong, episodic supernovae outflows. This causes DM cusp-to-core transformations for massive dwarfs ($M_{\rm vir} \sim 10^{10.5}$ M$_\odot$ or $M_\star/M_{\rm vir} \sim 10^{-3}$). Along with core formation, the overall density profile also becomes less concentrated. The APOSTLE and Auriga simulations, on the other hand, have relatively smooth and continuous star formation histories and therefore fewer intense episodes of supernovae feedback. The DM haloes remain cuspy throughout the mass range simulated (Bose et al., 2019). Cuspy, concentrated systems, once becoming satellites, are more resistant to tidal stripping. This is taken into consideration by the tidal evolution tracks described in Section B.2.6.

Therefore, as we can anticipate, an APOSTLE-like halo response would yield higher satellite counts than the more bursty NIHAO model. This is clearly shown by Figs. B.4 and B.5. We note that this effect is more pro-

nounced for massive satellites (as in Fig. B.5) than for the entire surviving population, which is dominated by low-mass systems (as in Fig. B.4). Specifically, the NIHAO emulator produces 20% fewer massive satellites than the APOSTLE emulator, while the difference in the abundance of all surviving satellites ($m > 10^6$ $M_\odot$) is only $\sim$ 7%. This is largely due to the fact that the two halo response relations mainly differ in the massive-dwarf regime, converging at the low mass end.

The relative importance of the halo response versus the baryonic disc of the host, in terms of its influence on satellite abundance, also depends on the model selection – for the whole population of surviving satellites, the disc effect is dominant, whereas for the massive dwarf subset, the disc effect is comparable to the halo response effect, both contributing to a $\sim 20 - 25\%$ difference.

Fig. B.5 shows that the halo-to-halo variance is dramatic, especially in the satellite spatial distributions. This highlights the importance of having a large sample if we hope to distinguish between feedback models. Hydro-simulation suites that consist of on the order of ten MW/M31 analogues would struggle in revealing the aforementioned differences (Samuel et al., 2020). Similarly, on the observational side, surveys of more MW/M31 analogues are needed. The SAGA survey (Geha et al., 2017), which will contain $\sim$100 MW-like systems when completed, will start to be a useful observational benchmark for differentiating feedback models based on the demographics of their satellite galaxies.

B.4.3 Effect of the disc potential

As we can expect, injecting a baryonic disc into the host galaxy has the effect of depleting satellites. This is simply because the disc is an extra source of tidal field and dynamical friction in addition to the smooth host halo. This satellite-depletion effect has been discussed by, e.g., Peñarrubia et al. (2010) and Garrison-Kimmel et al. (2017), using semi-analytical models and simulations. Here, we report consistent results. As shown in the right-hand panel of Fig. B.4, adding a disc reduces the abundance of surviving satellites by $\sim$20%. This effect is stronger towards the centre of the host and is not very sensitive to the halo response model.

In addition to depleting satellites, the disc also plays a secondary role of diversifying satellite structure. This is a subtle, but important, effect for reconciling the small-scale issues. Notably, the "too-big-to-fail" problem (TBTF) can be formulated as a tension between the narrow $v_{\rm max}$ distribution of subhaloes from ΛCDM models and the relatively broad $v_{\rm max}$ distribution of the observed massive satellites (e.g., Jiang & van den Bosch, 2015). The cusp-core issue is a tension that arises due to the fact that the observationally inferred DM inner slopes are quite diverse (e.g., Oman et al., 2015) whereas the ΛCDM subhalo inner slopes (in DM-only simulations) are almost exclusively cuspy. That is, both the TBTF and the cusp-core issues boil

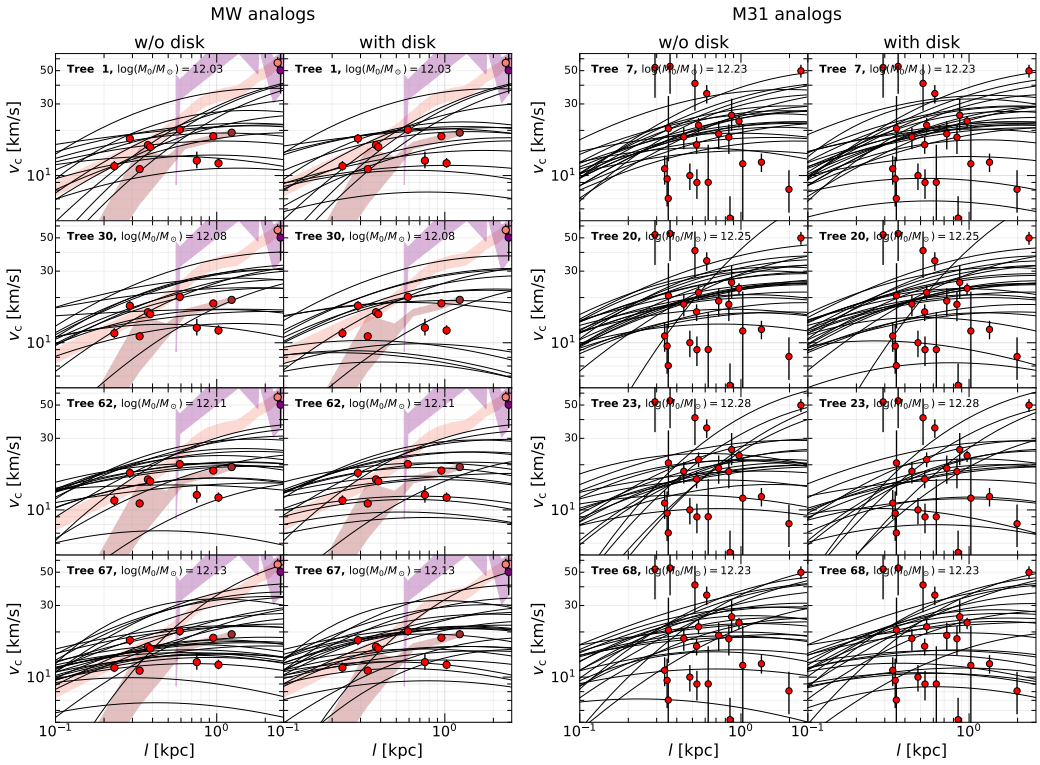

Figure B.6: Examples of rotation curves of massive satellites ($m_\star > 10^5$ M$_\odot$) of MW-sized ($M_{\rm vir} = 10^{12-12.15}$ M$_\odot$) and M31-sized ($M_{\rm vir} = 10^{12.15-12.3}$ M$_\odot$) host haloes at $z = 0$, from the NIHAO-emulating models. Each row is a random realization (indicated as "Tree *i*"), with the left-hand side and right-hand side panels having exactly the same merger history but differing in whether a baryonic disc is included (*right*) or ignored (*left*) when evolving the satellites (see Section B.4.1 for details about the disc setup). Symbols with error bars are kinematic data from the MW and M31 satellites compiled from the literature, where the red symbols are compiled by Garrison-Kimmel et al. (2019) using the references therein and the brown, pink, and purple symbols and the associated color bands are rotation curves of the Sagittarius dwarf, SMC, and LMC, respectively (Côté et al., 2000; Bekki & Stanimirović, 2009; van der Marel & Kallivayalil, 2014). Overall, the model rotation curves are in reasonable agreement with the observed kinematics, especially in the cases with a baryonic disc. The disc has a weak but noticeable effect of increasing the diversity of the rotation curves, as can be most clearly seen in `Tree 7`, `Tree 20`, and `Tree 30`.

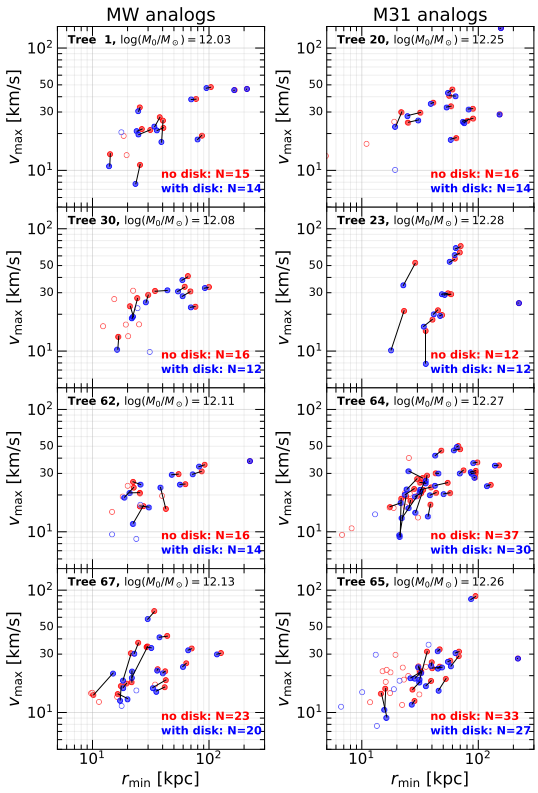

Figure B.7: Effect of a disc potential on satellite structure – $v_{\rm max}$ as a function of minimum galactocentric distance, $r_{\rm min}$, for surviving massive satellites ($m_\star > 10^5$ M$_\odot$) in MW-sized hosts ($M_{\rm vir} = 10^{12-12.15}$ M$_\odot$) and in M31-sized hosts ($M_{\rm vir} = 10^{12.15-12.3}$ M$_\odot$). Each panel shows the satellites in a *pair* of realizations with an identical, random merger tree evolved with and without a disc. Short black lines connect satellites shared in common (solid symbols) by the models with and without the disc, highlighting the change in $v_{\rm max}$. Open symbols represent the massive satellites that only belong to the disc models or the no-disc models. The numbers quoted in the lower right-hand corners of each panel are the numbers of surviving massive satellites. Focusing on the common satellites, we find that the disc generally decreases their $v_{\rm max}$ and $r_{\rm min}$. The $v_{\rm max}$ change is more pronounced for those satellites with smaller $r_{\rm min}$.

down to a structural diversity issue.

A commonly used diagnostic for TBTF is the comparison of the rotation curves (RCs) of massive satellites predicted by the model versus the circular velocities at certain radii observed for MW/M31 massive satellites, usually $v_{\rm circ}(l_{\rm eff})$. Fig. B.6 presents such examples from our NIHAO-emulating models. Overall, the agreement between the models and the data is decent, but we focus on comparing the results from the (merger tree-matched) models with and without the disc. We can see that the spread of the RCs is marginally larger in the models with a disc. This is especially clear in, e.g., `Tree 7`, `Tree 20`, and `Tree 30`. In the few cases, such as `Tree 67` and `Tree 68`, where the RCs in the no-disc models appear to be more scattered, the visual impression is actually misled by the fact that there are more satellites in the no-disc model. For an abundance-matched comparison, the RCs in the no-disc model are always more narrowly crowded and less diverse.

To better show the disc's role in broadening the structural diversity, we examine in Fig. B.7 the $v_{\rm max}$ change as a function of the minimum host-centric distance, $r_{\rm min}$, for individual massive ($m_\star > 10^5$ M$_\odot$) satellites in the merger tree-matched models with and without the disc. We can see that the disc decreases the $v_{\rm max}$ values by up to 50%, depending on $r_{\rm min}$. Generally, the closer a satellite gets to the host centre, the more that $v_{\rm max}$ decreases with respect to the no-disc case. The disc also marginally decreases the minimum galactocentric distances, as can be expected.

Fig. B.8 extends the analysis to the full ensemble, showing the median ratios of subhalo mass ($m_{\rm with\ disc}/m_{\rm no\ disc}$), maximum circular velocity ($m_{\rm with\ disc}/m_{\rm nodisc}$), subhalo concentration ($c_{2,\rm with\ disc}/c_{2,\rm no\ disc}$), and logarithmic inner density slope ($s_{38,\rm with\ disc}/s_{38,\rm no\ disc}$), as functions of the minimum host-centric distance measured in the simulations with disc, $r_{\rm min}$, of massive surviving satellites in all of the 100 realizations. Here, for the density slope we follow the convention in observational studies to measure it at fixed physical aperture (as opposed to a relative aperture of $0.01 l_{\rm vir}$ that is convenient for theoretical studies) – in particular, we use the average slope between $l = 0.3$ kpc and 0.8 kpc, $s_{38} \equiv -\ln[\rho(0.8\,{\rm kpc})/\rho(0.3\,{\rm kpc})]/\ln(0.8/0.3)$, following Relatores et al. (2019). On average, the disc decreases the subhalo mass by up to 60%, $v_{\rm max}$ by 20%, concentration by 5%, and steepens the density slope by 8%. Satellites need to reach small galactocentric distances to experience these changes: those not having been within 50 kpc of the galactic centre are barely affected.

We emphasize again that both the internal and external baryonic effects contribute a ~25% effect on the abundance and structure of satellite galaxies. The halo-to-halo variance due to different merging histories easily overwhelms these baryonic effects, unless large samples are utilized.

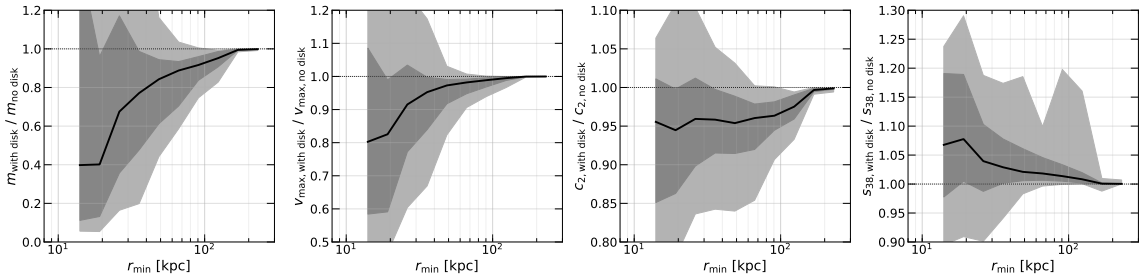

Figure B.8: The median ratios of subhalo mass, $v_{\rm max}$, concentration, and inner density slope (s_{38}, see Section B.4.3 for definition) between the models *with* and *without* the disc potential, all as a function of the minimum galactocentric distance (as measured in models with the disc), for all of the shared massive surviving satellites ($m_\star > 10^5$ M$_\odot$) in all of the 100 random realizations. Darker and lighter shaded bands indicate $16-84$ and $3-97$ percentiles, respectively. On average, the disc potential decreases satellite mass, $v_{\rm max}$, concentration, and increases the density slope – all in all, the disc increases satellite structural diversity.

B.5 Discussion: Survival versus disruption

It is natural to wonder what determines the fate of a satellite – under what internal and external conditions will a satellite survive, and under what conditions will it be disrupted? With the relatively large statistical samples provided by `SatGen`, we can address these questions quantitatively.

Fig. B.9 compares the distributions of massive surviving satellites ($m_\star > 10^5$ M$_\odot$) and of disrupted satellites ($m < 10^6$ M$_\odot$) in the space spanned by the minimum galactocentric distance ($r_{\rm min}$) versus virial mass at infall ($m_{\rm acc}$), concentration at infall ($c_{2,\rm acc}$), and logarithmic inner density slope at infall ($s_{0.01,\rm acc}$). In the first row of Fig. B.9, we include satellites accreted throughout cosmic history, whereas in the second and third rows of Fig. B.9, we consider satellites accreted at low redshift ($z_{\rm acc} < 1$) and higher redshift ($z_{\rm acc} = 1-2$), separately. We focus only on the NIHAO emulator results, but compare the models with and without the galactic disc potential.

There are several features worth mentioning. First, disruption occurs throughout the infall mass range. At the most massive end ($m_{\rm acc} \gtrsim 10^{11}$ M$_\odot$), disruption actually dominates over survival. This can be clearly seen via the $m_{\rm acc}$ distributions of satellites accreted after $z = 1$ (the top panel of the second row, first column, of Fig. B.9). This massive-end bump highlights the strong satellite mass dependence of dynamical friction: only massive satellites with $m/M \gtrsim 0.1$ undergo significant orbital decay. We caution that we have arbitrarily defined "disruption" as subhalo mass dropping below 10^6 M$_\odot$. This mass threshold is comparable or slightly better than the mass resolution of state-of-the-art zoom-in simulations of MW-sized haloes, where the DM particle mass is a few times 10^4 M$_\odot$ (e.g., Wetzel et al., 2016) and at least 100 particles are needed to resolve a substructure. Hence, our disruption threshold is comparable to that in high-resolution simulations. However, we emphasize that mass dropping below an arbitrary threshold does not necessarily correspond to physical disruption, and we refer interested readers to van den Bosch et al. (2018) for a

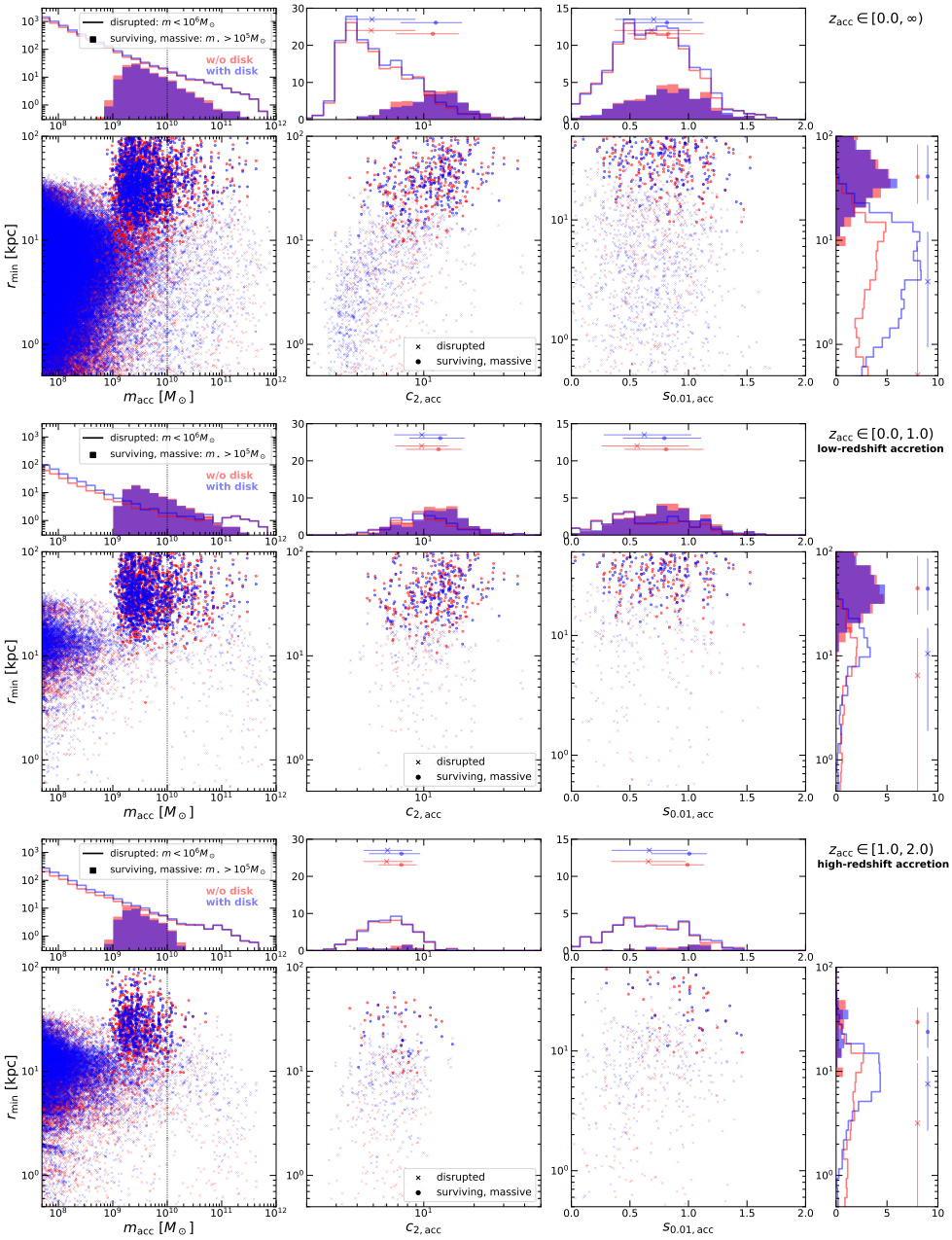

Figure B.9: Comparison of disrupted satellites ($m < 10^6\,\mathrm{M}_\odot$) and massive surviving satellites ($m_\star > 10^5\,\mathrm{M}_\odot$) in terms of their minimum host-centric distance versus mass, concentration, and inner slope at accretion, for the NIHAO-emulating models. The first row shows the results for satellites accreted throughout cosmic history. The second and third rows show results for satellites accreted at low redshift ($0 \leq z_{\rm acc} < 1$) and higher redshift ($1 \leq z_{\rm acc} < 2$), respectively. The top and side panels show the 1D marginalized histograms. Surviving satellites are shown as filled histograms while disrupted ones are shown as empty steps. The middle column ($r_{\rm min}$ versus $c_{2,\rm acc}$) and right-hand column ($r_{\rm min}$ versus $s_{0.01,\rm acc}$) focus only on satellites with $m_{\rm acc} > 10^{10}\,\mathrm{M}_\odot$. Key takeaways: (1) Disruption occurs throughout the mass range, with a hump at the massive end, illustrating that massive satellites experience stronger dynamical friction. (2) Surviving satellites have higher concentration and cuspier density profiles at infall. However, the concentration trend largely reflects a progenitor bias (namely that concentration anti-correlates with redshift) and is significantly reduced if focusing on satellites accreted in the same redshift range. (3) The disc potential causes disruption to occur at larger galactocentric distances.

thorough discussion.

Second, surviving satellites were more concentrated and more cuspy at accretion. Specifically, if we focus on massive satellites with $m_{\rm acc} > 10^{10}\,{\rm M}_\odot$, the surviving ones have a median concentration of $c_{2,\rm acc} \approx 11$ and a median inner slope of $s_{0.01,\rm acc} \approx 0.8$, while the disrupted ones have a median concentration of $c_{2,\rm acc} \approx 5$ and a median slope of $s_{0.01,\rm acc} \approx 0.7$. At face value, the concentration trend seems to have a simple interpretation: denser haloes are more resistant to tidal disruption. While this statement is true on its own, it is actually not the main factor at play here. The time spent in the host halo is more important for the disruption of a subhalo than properties of the initial density profile. This can be seen from the second and third rows of Fig. B.9: selecting satellites by infall redshift significantly reduces the difference in $c_{2,\rm acc}$ between the disrupted and surviving populations. Halo concentration at fixed mass anti-correlates with redshift (e.g., Dutton & Macciò, 2014), so the satellites that were accreted earlier (and thus exposed for a longer time to the tidal field of the host) naturally tend to have lower concentrations. However, the inner cuspiness is almost independent of redshift. In fact, taking $z_{\rm acc}$ bins makes the slope difference more pronounced: for $z_{\rm acc} \in [1, 2)$, the surviving satellites have $s_{0.01,\rm acc} \approx 1$, and the disrupted ones have $s_{0.01,\rm acc} \approx 0.6$.

Third, the disc significantly changes the minimum galactocentric distance at which disruption takes place. In particular, without a disc potential, satellites can travel to as close as $r_{\rm min} \lesssim 1\,{\rm kpc}$ from the galactic centre before becoming disrupted, whereas with a disc, most disruption events occur outside 1 kpc, with a median $r_{\rm min}$ of 4 kpc. This again illustrates the disruptive role of the galactic disc. Massive surviving satellites can seldom travel within 10 kpc of the galactic centre. In this way, the Solar neighbourhood is shielded against massive satellites.

B.6 Conclusion

In this chapter, we presented a new semi-analytical model (`SatGen`) for generating satellite galaxy populations. The model is devised to generate statistical samples of satellite galaxy populations for desired host properties, emulating zoom-in cosmological simulations and outperforming simulations in statistical power. It combines halo merger trees, empirical relations that describe the galaxy-halo connection, and analytical prescriptions for satellite evolution, incorporating new developments in these areas. Its improvements and features can be summarized as follows:

- It uses the Parkinson et al. (2008) algorithm to generate halo merger trees, with parameters recently re-calibrated by Benson (2017). It can also be applied to merger trees from N-body simulations.

- It supports halo density profiles that are more flexible than the NFW profile, including the Einasto

profile and the Dekel+ profile, the latter of which has useful analytical properties. It also uses the MN profile for describing discs.

- It can be used to emulate hydro-simulations with different sub-grid baryonic physics via an empirical treatment of the halo response to star formation and feedback, as extracted from zoom-in hydro-simulations of field galaxies.

- It makes use of stellar-mass-halo-mass relations from halo abundance matching, as well as galaxy-size-halo-size relations extracted from hydro-simulations, in order to initialize the baryonic properties.

- It supports satellite orbit integration in *composite* host potentials, consisting of (combinations of) a DM halo, baryonic disc, and stellar bulge.

- It uses tidal evolution tracks obtained from high-resolution idealized simulations from Peñarrubia et al. (2008); Peñarrubia et al. (2010) and Errani et al. (2015, 2018), following the structural evolution of satellites. This, together with the halo response relations, enables `SatGen` to propagate the baryonic effects seen in hydro-simulations to the satellite populations – a task that is difficult for simulations because of the high numerical resolution required.

We presented a proof-of-concept application of `SatGen`. We generated samples much larger than state-of-the-art zoom-in simulations for MW and M31 at comparable numerical resolution. We experimented with different halo response models, using `SatGen` to emulate simulations with bursty star formation and strong feedback (e.g., NIHAO and FIRE) and simulations with smoother star formation, and thus negligible halo response, in massive dwarfs (e.g., APOSTLE and Auriga). We also experimented with models with and without a galactic disc potential in order to quantify the influence of the disc on satellite statistics. In other words, we explored the internal (halo response) and external (host-disc) baryonic effects on satellite properties. The conclusions of this study are as follows:

- We find that the model predictions of the $v_{\rm max}$ function, rotation curves, and spatial distributions of bright satellites with $m_\star > 10^5\,{\rm M}_\odot$ are in good agreement with observations. This is achieved without fine-tuning model parameters.

- Different halo response models yield slightly different satellite abundances: on average, the NIHAO emulator yields 25% less satellites with $m_\star > 10^5\ {\rm M}_\odot$ within 300 kpc of the galactic centre than the APOSTLE emulator. The effect is smaller if we include all of the surviving satellites, illustrating the fact that the difference in the halo response is most prominent for massive dwarfs. Given the

large halo-to-halo variance as revealed by the model, and given the limited observational sample, it currently remains difficult to use the observed satellite spatial distribution to distinguish between the two feedback patterns.

- Adding a disc potential to the host causes, on average, a 20% (30%) reduction in satellite number count within 300 (100) kpc. In addition to satellite depletion, the disc slightly increases the structural diversity of massive satellite dwarfs. On average, a disc decreases the satellite $v_{\rm max}$ by up to 20%, concentration by up to 5%, and increase the density slope measured at the fixed physical aperture of $0.3 - 0.8$ kpc by up to 8%, depending on the minimum galactocentric distance that the satellite can reach. This helps with alleviating the small-scale problems of ΛCDM.

- The fate of a massive satellite galaxy ($m_{\rm acc} > 10^{10}$ M$_\odot$) depends on how close it gets to the galactic centre: the surviving satellites seldom reach within 10 kpc of the centre, whereas the disrupted ones have a minimum galactocentric distance of $r_{\rm min} \sim 4$ kpc (or $\lesssim 1$ kpc if there was no galactic disc). The fate also depends on the initial structure at infall: more concentrated and cuspier haloes are more likely to survive. However, the concentration trend is largely due to a progenitor bias, in the sense that satellites that have been exposed to the tidal field for a longer time, i.e., those that were accreted earlier, have lower concentration at accretion because of the anti-correlation between halo concentration and redshift.

Overall, we have shown that `SatGen` can emulate numerical simulations of very high resolution decently, capturing the bulk of the baryonic effects on the abundance, spatial distribution, and internal structure of satellites. Thanks to the tidal evolution recipes that are extracted from high-resolution idealized simulations, it avoids the numerical artifacts of over-stripping. Simulating a statistically large sample of MW/M31-sized systems, not to mention galaxy groups or clusters, while retaining the resolution for satellite dwarfs is computationally challenging for numerical simulations. Therefore, the `SatGen` model complements simulations nicely in terms of statistical power. In an upcoming work (Jiang et al., in prep), we use `SatGen` to study satellites of group-sized hosts and explore the conditions for forming ultra-diffuse galaxies and compact dwarf satellites. The `SatGen` code is made publicly available at https://github.com/shergreen/SatGen.

Bibliography

[1] Abazajian K., et al., 2019, arXiv e-prints, p. arXiv:1907.04473

[2] Abbott T. M. C., et al., 2018, Phys. Rev. D, 98, 043526

[3] Abraham R. G., Tanvir N. R., Santiago B. X., Ellis R. S., Glazebrook K., van den Bergh S., 1996, MNRAS, 279, L47

[4] Ackermann M., et al., 2014, ApJ, 787, 18

[5] Ade P., et al., 2019, J. Cosmology Astropart. Phys., 2019, 056

[6] Adhikari S., Dalal N., Chamberlain R. T., 2014, J. Cosmology Astropart. Phys., 2014, 019

[7] Aguilar L. A., White S. D. M., 1986, ApJ, 307, 97

[8] Agustsson I., Brainerd T. G., 2006, ApJ, 644, L25

[9] Alcock C., et al., 2000, ApJ, 542, 281

[10] Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409

[11] Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781

[12] Ameglio S., Borgani S., Pierpaoli E., Dolag K., Ettori S., Morandi A., 2009, MNRAS, 394, 479

[13] Anbajagane D., Evrard A. E., Farahi A., Barnes D. J., Dolag K., McCarthy I. G., Nelson D., Pillepich A., 2020, MNRAS, 495, 686

[14] Ando S., Ishiyama T., Hiroshima N., 2019, Galaxies, 7, 68

[15] Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046

[16] Ansarifard S., et al., 2020, A&A, 634, A113

[17] Antonini F., Merritt D., 2012, ApJ, 745, 83

[18] Applegate D. E., et al., 2014, MNRAS, 439, 48

[19] Applegate D. E., et al., 2016, MNRAS, 457, 1522

[20] Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., van der Hulst J. M., 2007, ApJ, 655, L5

[21] Armitage T. J., Kay S. T., Barnes D. J., 2019, MNRAS, 484, 1526

[22] Arnaud K. A., 1996, in Jacoby G. H., Barnes J., eds, Astronomical Society of the Pacific Conference Series

Vol. 101, Astronomical Data Analysis Software and Systems V. p. 17

[23] Arnaud M., Pratt G. W., Piffaretti R., Böhringer H., Croston J. H., Pointecouteau E., 2010, A&A, 517, A92

[24] Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376

[25] Aung H., Nagai D., Rozo E., García R., 2021, MNRAS, 502, 1041

[26] Avestruz C., Nagai D., Lau E. T., 2016, ApJ, 833, 227

[27] Azzaro M., Patiri S. G., Prada F., Zentner A. R., 2007, MNRAS, 376, L43

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

[29] Banik U., van den Bosch F. C., 2021, ApJ, 912, 43

[30] Banik N., Bovy J., Bertone G., Erkal D., de Boer T. J. L., 2019, arXiv e-prints, p. arXiv:1911.02663

[31] Barnes J., Hut P., 1986, Nature, 324, 446

[32] Barnes J., White S. D. M., 1984, MNRAS, 211, 753

[33] Barnes D. J., Kay S. T., Henson M. A., McCarthy I. G., Schaye J., Jenkins A., 2017a, MNRAS, 465, 213

[34] Barnes D. J., et al., 2017b, MNRAS, 471, 1088

[35] Barnes D. J., Vogelsberger M., Pearce F. A., Pop A.-R., Kannan R., Cao K., Kay S. T., Hernquist L., 2021, MNRAS,

[36] Barret D., et al., 2016, in Proc. SPIE. p. 99052F, doi:10.1117/12.2232432

[37] Bartels R., Ando S., 2015, Phys. Rev. D, 92, 123508

[38] Basu K., et al., 2021, Experimental Astronomy,

[39] Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., Sijacki D., 2010, ApJ, 725, 91

[40] Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012, ApJ, 758, 74

[41] Battaglia N., Hill J. C., Murray N., 2015, ApJ, 812, 154

[42] Baxter E., et al., 2017, ApJ, 841, 18

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

[44] Behroozi P. S., Wechsler R. H., Conroy C., 2013b, ApJ, 770, 57

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

[46] Bekki K., Stanimirović S., 2009, MNRAS, 395, 342

[47] Benson A. J., 2017, MNRAS, 467, 3454

[48] Benson A. J., 2020, MNRAS, 493, 1268

[49] Benson A. J., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2001, MNRAS, 327, 1041

[50] Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002a, MNRAS, 333, 156

[51] Benson A. J., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2002b, MNRAS, 333, 177

[52] Bergström L., Edsjö J., Gondolo P., Ullio P., 1999, Phys. Rev. D, 59, 043506

[53] Berlind A. A., et al., 2003, ApJ, 593, 1

[54] Bernardeau F., 1994, ApJ, 433, 1

[55] Biffi V., Dolag K., Böhringer H., Lemson G., 2012, MNRAS, 420, 3545

[56] Biffi V., Dolag K., Böhringer H., 2013, MNRAS, 428, 1395

[57] Biffi V., et al., 2016, ApJ, 827, 112

[58] Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press

[59] Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, ApJ, 301, 27

[60] Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361

[61] Bolliet B., Comis B., Komatsu E., Macías-Pérez J. F., 2018, MNRAS, 477, 4957

[62] Bonaca A., et al., 2020, ApJ, 892, L37

[63] Bonamente M., Joy M., LaRoque S. J., Carlstrom J. E., Nagai D., Marrone D. P., 2008, ApJ, 675, 106

[64] Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

[65] Borm K., Reiprich T. H., Mohammed I., Lovisari L., 2014, A&A, 567, A65

[66] Bose S., et al., 2017, MNRAS, 464, 4520

[67] Bose S., et al., 2019, MNRAS, 486, 4790

[68] Boyarsky A., Drewes M., Lasserre T., Mertens S., Ruchayskiy O., 2019, Prog. Part. Nucl. Phys., 104, 1

[69] Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93

[70] Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, MNRAS, 406, 896

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

[72] Brainerd T. G., 2005, ApJ, 628, L101

[73] Breiman L., 2001, Mach. Learn., 45, 5

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

[75] Browne M. W., 2000, J. Math. Psychol., 44, 108

[76] Bryan G. L., Norman M. L., 1998, ApJ, 495, 80

[77] Bullock J. S., Boylan-Kolchin M., 2017, preprint, (arXiv:1707.04256)

[78] Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559

[79] Buote D. A., Tsai J. C., 1995, ApJ, 452, 522

[80] Burkert A., 1995, ApJ, 447, L25

[81] Burkert A., 2000, ApJ, 534, L143

[82] Burkert A., 2020, ApJ, 904, 161

[83] Busch P., White S. D. M., 2017, MNRAS, 470, 4767

[84] Calderon V. F., Berlind A. A., 2019, MNRAS, 490, 2367

[85] Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villar-

real A., 2018, MNRAS, 477, 359

[86] Carlberg R. G., 1994, ApJ, 433, 468

[87] Carlberg R. G., 2012, ApJ, 748, 20

[88] Carleton T., Errani R., Cooper M., Kaplinghat M., Peñarrubia J., Guo Y., 2019, MNRAS, 485, 382

[89] Carlsten S. G., Greene J. E., Peter A. H. G., Greco J. P., Beaton R. L., 2020, ApJ, 902, 124

[90] Chadayammuri U., Tremmel M., Nagai D., Babul A., Quinn T., 2020, arXiv e-prints, p. arXiv:2001.06532

[91] Chandrasekhar S., 1943, ApJ, 97, 255

[92] Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, MNRAS, 460, 3100

[93] Chen H., Avestruz C., Kravtsov A. V., Lau E. T., Nagai D., 2019, MNRAS, 490, 2380

[94] Chluba J., Nagai D., Sazonov S., Nelson K., 2012, MNRAS, 426, 510

[95] Chluba J., Switzer E., Nelson K., Nagai D., 2013, MNRAS, 430, 3054

[96] Ciotti L., Pellegrini S., 1996, MNRAS, 279, 240

[97] Clerc N., et al., 2018, A&A, 617, A92

[98] Cohn J. D., Battaglia N., 2020, MNRAS, 491, 1575

[99] Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

[100] Colín P., Avila-Reese V., González-Samaniego A., Velázquez H., 2015, ApJ, 803, 28

[101] Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720

[102] Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

[103] Conselice C. J., 2003, ApJS, 147, 1

[104] Corless V. L., King L. J., 2007, MNRAS, 380, 149

[105] Côté S., Carignan C., Freeman K. C., 2000, AJ, 120, 3027

[106] Courteau S., Rix H.-W., 1999, ApJ, 513, 561

[107] Courteau S., Dutton A. A., van den Bosch F. C., MacArthur L. A., Dekel A., McIntosh D. H., Dale D. A., 2007, ApJ, 671, 203

[108] Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303

[109] Czakon N. G., et al., 2015, ApJ, 806, 18

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

[111] Dalal N., Kochanek C. S., 2002, ApJ, 572, 25

[112] Dekel A., Devor J., Hetzroni G., 2003, MNRAS, 341, 326

[113] Dekel A., Ishai G., Dutton A. A., Maccio A. V., 2017, MNRAS, 468, 1005

[114] Delos M. S., 2019, Phys. Rev. D, 100, 063505

[115] Di Cintio A., Brook C. B., Macciò A. V., Stinson G. S., Knebe A., Dutton A. A., Wadsley J., 2014a, MNRAS,

437, 415

[116] Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Stinson G. S., Knebe A., 2014b, MNRAS, 441, 2986

[117] Diemand J., Moore B., Stadel J., 2004, MNRAS, 352, 535

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

[119] Diemand J., Kuhlen M., Madau P., Zemp M., Moore B., Potter D., Stadel J., 2008, Nature, 454, 735

[120] Diemer B., 2017, ApJS, 231, 5

[121] Diemer B., 2018, ApJS, 239, 35

[122] Diemer B., 2020, ApJS, 251, 17

[123] Diemer B., 2021, ApJ, 909, 112

[124] Diemer B., Kravtsov A. V., 2014, ApJ, 789, 1

[125] Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108

[126] Diemer B., Mansfield P., Kravtsov A. V., More S., 2017, ApJ, 843, 140

[127] Dietrich J. P., et al., 2019, MNRAS, 483, 2871

[128] Dodelson S., 2003, Modern cosmology. Academic Press

[129] Dolag K., Schindler S., 2000, A&A, 364, 491

[130] Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

[131] Dolag K., Gaensler B. M., Beck A. M., Beck M. C., 2015, MNRAS, 451, 4277

[132] Dolag K., Komatsu E., Sunyaev R., 2016, MNRAS, 463, 1797

[133] Drakos N. E., Taylor J. E., Benson A. J., 2017, MNRAS, 468, 2345

[134] Drakos N. E., Taylor J. E., Benson A. J., 2020, MNRAS, 494, 378

[135] Drinkwater M. J., Gregg M. D., Hilker M., Bekki K., Couch W. J., Ferguson H. C., Jones J. B., Phillipps S., 2003, Nature, 423, 519

[136] Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

[137] Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007, ApJ, 654, 27

[138] Dutton A. A., et al., 2016, MNRAS, 461, 2658

[139] Dutton A. A., Macciò A. V., Buck T., Dixon K. L., Blank M., Obreja A., 2019, MNRAS, 486, 655

[140] Eckert D., Molendi S., Paltani S., 2011, A&A, 526, A79

[141] Eckert D., Ettori S., Molendi S., Vazza F., Paltani S., 2013, A&A, 551, A23

[142] Eckert D., Roncarelli M., Ettori S., Molendi S., Vazza F., Gastaldello F., Rossetti M., 2015, MNRAS, 447, 2198

[143] Eckert D., et al., 2019, A&A, 621, A40

[144] Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705

[145] Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

[146] El-Zant A., Shlosman I., Hoffman Y., 2001a, ApJ, 560, 636

[147] El-Zant A., Shlosman I., Hoffman Y., 2001b, ApJ, 560, 636

[148] Erkal D., Belokurov V., Bovy J., Sanders J. L., 2016, MNRAS, 463, 102

[149] Errani R., Navarro J. F., 2021, MNRAS, 505, 18

[150] Errani R., Peñarrubia J., 2020, MNRAS, 491, 4591

[151] Errani R., Peñarrubia J., Tormen G., 2015, MNRAS, 449, L46

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

[153] Errani R., Peñarrubia J., Walker M. G., 2018, MNRAS, 481, 5073

[154] Ettori S., Dolag K., Borgani S., Murante G., 2006, MNRAS, 365, 1021

[155] Ettori S., et al., 2019, A&A, 621, A39

[156] Evrard A. E., 1990, ApJ, 363, 349

[157] Facchinetti G., Lavalle J., Stref M., 2020, arXiv e-prints, p. arXiv:2007.10392

[158] Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267

[159] Faltenbacher A., Jing Y. P., Li C., Mao S., Mo H. J., Pasquali A., van den Bosch F. C., 2008, ApJ, 675, 146

[160] Farahi A., Evrard A. E., McCarthy I., Barnes D. J., Kay S. T., 2018, MNRAS, 478, 2618

[161] Fellhauer M., Lin D. N. C., 2007, MNRAS, 375, 604

[162] Flender S., Nagai D., McDonald M., 2017, ApJ, 837, 124

[163] Fong M., Han J., 2021, MNRAS, 503, 4250

[164] Freund Y., Schapire R. E., 1996, in Proceedings of the 13th International Conference on Machine Learning. Morgan Kaufmann, pp 148–156

[165] Freundlich J., Dekel A., Jiang F., Ishai G., Cornuault N., Lapiner S., Dutton A. A., Macciò A. V., 2020a, MNRAS, 491, 4523

[166] Freundlich J., et al., 2020b, MNRAS, 499, 2912

[167] Friedman J. H., 2001, Ann. Statist., 29, 1189

[168] Fujii M., Funato Y., Makino J., 2006, PASJ, 58, 743

[169] Fujita Y., Aung H., 2019, ApJ, 875, 26

[170] Gaia Collaboration et al., 2016, A&A, 595, A1

[171] Gan J., Kang X., van den Bosch F. C., Hou J., 2010, MNRAS, 408, 2201

[172] Gao F., Han L., 2012, Comput. Optim. Appl., 51, 259

[173] Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004, MNRAS, 355, 819

[174] Gao L., Navarro J. F., Frenk C. S., Jenkins A., Springel V., White S. D. M., 2012, MNRAS, 425, 2169

[175] Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014a, MNRAS, 438, 2578

[176] Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Kirby E. N., 2014b, MNRAS, 444, 222

[177] Garrison-Kimmel S., et al., 2017, MNRAS, 471, 1709

[178] Garrison-Kimmel S., et al., 2019, MNRAS, 487, 1380

[179] Gaskin J. A., et al., 2019, J. Astron. Telesc. Instrum. Syst., 5, 021001

[180] Geha M., et al., 2017, ApJ, 847, 4

[181] Gentile G., Salucci P., Klein U., Vergani D., Kalberla P., 2004, MNRAS, 351, 903

[182] Géron A., 2017, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st edn. O'Reilly Media, Inc.

[183] Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146

[184] Ghirardini V., et al., 2019, A&A, 621, A41

[185] Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399

[186] Gilman D., Birrer S., Nierenberg A., Treu T., Du X., Benson A., 2020a, MNRAS, 491, 6077

[187] Gilman D., Birrer S., Treu T., 2020b, A&A, 642, A194

[188] Giocoli C., Tormen G., Sheth R. K., van den Bosch F. C., 2010, MNRAS, 404, 502

[189] Gnedin O. Y., Ostriker J. P., 1999, ApJ, 513, 626

[190] Gnedin O. Y., Hernquist L., Ostriker J. P., 1999a, ApJ, 514, 109

[191] Gnedin O. Y., Lee H. M., Ostriker J. P., 1999b, ApJ, 522, 935

[192] Golse G., Kneib J. P., 2002, A&A, 390, 821

[193] Gonzalez-Casado G., Mamon G. A., Salvador-Sole E., 1994, ApJ, 433, L61

[194] Grand R. J. J., et al., 2017, MNRAS, 467, 179

[195] Green S. B., van den Bosch F. C., 2019, MNRAS, 490, 2091

[196] Green S. B., Ntampaka M., Nagai D., Lovisari L., Dolag K., Eckert D., ZuHone J. A., 2019, ApJ, 884, 33

[197] Green S. B., Aung H., Nagai D., van den Bosch F. C., 2020, MNRAS, 496, 2743

[198] Green S. B., van den Bosch F. C., Jiang F., 2021a, MNRAS, 503, 4075

[199] Green S. B., van den Bosch F. C., Jiang F., 2021b, MNRAS, arXiv, 2110.13044

[200] Griffen B. F., Ji A. P., Dooley G. A., Gómez F. A., Vogelsberger M., O'Shea B. W., Frebel A., 2016, ApJ, 818, 10

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

[202] Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111

[203] Guth A. H., Pi S. Y., 1982, Phys. Rev. Lett., 49, 1110

[204] Hahn O., Martizzi D., Wu H.-Y., Evrard A. E., Teyssier R., Wechsler R. H., 2017, MNRAS, 470, 166

[205] Hallman E. J., O'Shea B. W., Burns J. O., Norman M. L., Harkness R., Wagner R., 2007, ApJ, 671, 27

[206] Hambrick D. C., Ostriker J. P., Johansson P. H., Naab T., 2011, MNRAS, 413, 2421

[207] Han J., Cole S., Frenk C. S., Jing Y., 2016, MNRAS, 457, 1208

[208] Hayashi K., Chiba M., 2015, ApJ, 810, 22

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

[210] Hayashi E., Navarro J. F., Springel V., 2007, MNRAS, 377, 50

[211] Hayashi K., Ichikawa K., Matsumoto S., Ibe M., Ishigaki M. N., Sugai H., 2016, MNRAS, 461, 2914

[212] Hearin A. P., Zentner A. R., Berlind A. A., Newman J. A., 2013, MNRAS, 433, 659

[213] Henden N. A., Puchwein E., Shen S., Sijacki D., 2018, MNRAS, 479, 5385

[214] Henden N. A., Puchwein E., Sijacki D., 2019, MNRAS, 489, 2439

[215] Henson M. A., Barnes D. J., Kay S. T., McCarthy I. G., Schaye J., 2017, MNRAS, 465, 3361

[216] Hezaveh Y. D., et al., 2016, ApJ, 823, 37

[217] Hildebrandt H., et al., 2020, A&A, 633, A69

[218] Hiroshima N., Ando S., Ishiyama T., 2018, Phys. Rev. D, 97, 123002

[219] Hitomi Collaboration et al., 2018, PASJ, 70, 9

[220] Ho M., Rau M. M., Ntampaka M., Farahi A., Trac H., Póczos B., 2019, ApJ, 887, 25

[221] Hoekstra H., Mahdavi A., Babul A., Bildfell C., 2012, MNRAS, 427, 1298

[222] Hoekstra H., Herbonnet R., Muzzin A., Babul A., Mahdavi A., Viola M., Cacciato M., 2015, MNRAS, 449, 685

[223] Hoerl A. E., Kennard R. W., 1970, Technometrics, 12, 55

[224] Hojjati A., et al., 2017, MNRAS, 471, 1565

[225] Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

[226] Hopkins P. F., et al., 2018, MNRAS, 480, 800

[227] Hui L., Ostriker J. P., Tremaine S., Witten E., 2017, Phys. Rev. D, 95, 043541

[228] Hurier G., Angulo R. E., 2018, A&A, 610, L4

[229] Iapichino L., Brüggen M., 2012, MNRAS, 423, 2781

[230] Inogamov N. A., Sunyaev R. A., 2003, Astronomy Letters, 29, 791

[231] Inoue S., Saitoh T. R., 2011, MNRAS, 418, 2527

[232] Israel H., Schellenberger G., Nevalainen J., Massey R., Reiprich T. H., 2015, MNRAS, 448, 814

[233] Jackson R. A., et al., 2021, MNRAS, 502, 1785

[234] Jiang F., 2016, PhD thesis, Yale University, https://www.proquest.com/dissertations-theses/characterizing-substructure-dark-matter-halos/docview/1923431768/se-2?accountid=15172

[235] Jiang F., van den Bosch F. C., 2014, MNRAS, 440, 193

[236] Jiang F., van den Bosch F. C., 2015, MNRAS, 453, 3575

[237] Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848

[238] Jiang F., van den Bosch F. C., 2017, MNRAS, 472, 657

[239] Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, ApJ, 675, 1095

[240] Jiang L., Cole S., Sawala T., Frenk C. S., 2015, MNRAS, 448, 1674

[241] Jiang F., Dekel A., Freundlich J., Romanowsky A. J., Dutton A. A., Macciò A. V., Di Cintio A., 2019a, MNRAS, 487, 5272

[242] Jiang F., et al., 2019b, MNRAS, 488, 4801

[243] Jiang F., Dekel A., Freundlich J., van den Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2021, MNRAS, 502, 621

[244] Jing Y. P., Suto Y., 2002, ApJ, 574, 538

[245] Joudaki S., et al., 2020, A&A, 638, L1

[246] Jungman G., Kamionkowski M., Griest K., 1996, Phys. Rep., 267, 195

[247] Käfer F., Finoguenov A., Eckert D., Clerc N., Ramos-Ceja M. E., Sanders J. S., Ghirardini V., 2020, A&A, 634, A8

[248] Kaiser N., 1986, MNRAS, 222, 323

[249] Kale R., Venturi T., Giacintucci S., Dallacasa D., Cassano R., Brunetti G., Macario G., Athreya R., 2013, A&A, 557, A99

[250] Kampakoglou M., Benson A. J., 2007, MNRAS, 374, 775

[251] Kang X., van den Bosch F. C., Yang X., Mao S., Mo H. J., Li C., Jing Y. P., 2007, MNRAS, 378, 1531

[252] Kaplinghat M., Tulin S., Yu H.-B., 2016, Phys. Rev. Lett., 116, 041302

[253] Katz N., White S. D. M., 1993, ApJ, 412, 455

[254] Kay S. T., Peel M. W., Short C. J., Thomas P. A., Young O. E., Battye R. A., Liddle A. R., Pearce F. R., 2012, MNRAS, 422, 1999

[255] Kazantzidis S., Zentner A. R., Kravtsov A. V., 2006, ApJ, 641, 647

[256] Keeton C. R., Moustakas L. A., 2009, ApJ, 699, 1720

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

[258] Khedekar S., Churazov E., Kravtsov A., Zhuravleva I., Lau E. T., Nagai D., Sunyaev R., 2013, MNRAS, 431, 954

[259] Khochfar S., Burkert A., 2006, A&A, 445, 403

[260] Kim S. Y., Peter A. H. G., Hargis J. R., 2018, Phys. Rev. Lett., 121, 211302

[261] King I., 1962, AJ, 67, 471

[262] Kitzbichler M. G., White S. D. M., 2008, MNRAS, 391, 1489

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

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

[265] Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

[266] Klypin A., Prada F., Yepes G., Heß S., Gottlöber S., 2015, MNRAS, 447, 3693

[267] Knebe A., Gill S. P. D., Gibson B. K., Lewis G. F., Ibata R. A., Dopita M. A., 2004, ApJ, 603, 7

[268] Knebe A., Arnold B., Power C., Gibson B. K., 2008, MNRAS, 386, 1029

[269] Knebe A., et al., 2011, MNRAS, 415, 2293

[270] Knebe A., et al., 2013, MNRAS, 435, 1618

[271] Komatsu E., Seljak U., 2001, MNRAS, 327, 1353

[272] Komatsu E., Seljak U., 2002, MNRAS, 336, 1256

[273] Komatsu E., et al., 2011, ApJS, 192, 18

[274] Krause E., Pierpaoli E., Dolag K., Borgani S., 2012, MNRAS, 419, 1766

[275] Kravtsov A. V., 2013, ApJ, 764, L31

[276] Kravtsov A. V., Borgani S., 2012, ARA&A, 50, 353

[277] Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

[278] Kravtsov A. V., Nagai D., Vikhlinin A. A., 2005, ApJ, 625, 588

[279] Kravtsov A. V., Vikhlinin A., Nagai D., 2006, ApJ, 650, 128

[280] Kuijken K., Dubinski J., 1994, MNRAS, 269, 13

[281] Lacey C., Cole S., 1993, MNRAS, 262, 627

[282] Lange J. U., van den Bosch F. C., Zentner A. R., Wang K., Villarreal A. S., 2019, MNRAS, 487, 3112

[283] Lau E. T., Kravtsov A. V., Nagai D., 2009, ApJ, 705, 1129

[284] Lau E. T., Nagai D., Nelson K., 2013, ApJ, 777, 151

[285] Lau E. T., Nagai D., Avestruz C., Nelson K., Vikhlinin A., 2015, ApJ, 806, 68

[286] Law D. R., Majewski S. R., 2010, ApJ, 714, 229

[287] Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2017, MNRAS, 466, 4442

[288] Lee E., Chluba J., Kay S. T., Barnes D. J., 2020, MNRAS, 493, 3274

[289] Li Y., Mo H. J., Gao L., 2008, MNRAS, 389, 1419

[290] Li Z.-Z., Zhao D.-H., Jing Y. P., Han J., Dong F.-Y., 2020, ApJ, 905, 177

[291] Libeskind N. I., Frenk C. S., Cole S., Helly J. C., Jenkins A., Navarro J. F., Power C., 2005, MNRAS, 363, 146

[292] Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013, MNRAS, 428, 2489

[293] Łokas E. L., Mamon G. A., 2001, MNRAS, 321, 155

[294] Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163

[295] Lovell M. R., et al., 2012, MNRAS, 420, 2318

[296] Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300

[297] Lovisari L., et al., 2017, ApJ, 846, 51

[298] Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, ApJ, 692, 931

[299] Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, MNRAS, 460, 1214

[300] Ludlow A. D., Schaye J., Bower R., 2019, MNRAS, 488, 3663

[301] Mahdavi A., Hoekstra H., Babul A., Bildfell C., Jeltema T., Henry J. P., 2013, ApJ, 767, 116

[302] Makiya R., Ando S., Komatsu E., 2018, MNRAS, 480, 3928

[303] Makiya R., Hikage C., Komatsu E., 2020, PASJ, 72, 26

[304] Man Z.-Y., Peng Y.-J., Shi J.-J., Kong X., Zhang C.-P., Dou J., Guo K.-X., 2019, ApJ, 881, 74

[305] Mansfield P., Avestruz C., 2020, MNRAS, 500, 3309

[306] Mansfield P., Kravtsov A. V., Diemer B., 2017, ApJ, 841, 34

[307] Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., von der Linden A., Schmidt R. W., 2014, MNRAS, 440, 2077

[308] Mantz A. B., et al., 2015a, MNRAS, 446, 2205

[309] Mantz A. B., Allen S. W., Morris R. G., Schmidt R. W., von der Linden A., Urban O., 2015b, MNRAS, 449, 199

[310] Mantz A. B., Allen S. W., Morris R. G., von der Linden A., 2018, MNRAS, 473, 3072

[311] Mao Y.-Y., Williamson M., Wechsler R. H., 2015, ApJ, 810, 21

[312] Marrone D. P., et al., 2012, ApJ, 754, 119

[313] Mather J. C., et al., 1990, ApJ, 354, L37

[314] Maughan B. J., 2007, ApJ, 668, 772

[315] McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936

[316] McConnachie A. W., 2012, AJ, 144, 4

[317] McCourt M., Quataert E., Parrish I. J., 2013, MNRAS, 432, 404

[318] Medezinski E., et al., 2018, PASJ, 70, S28

[319] Meneghetti M., et al., 2020, Science, 369, 1347

[320] Merloni A., et al., 2012, arXiv e-prints, p. arXiv:1209.3114

[321] Milgrom M., 1983, ApJ, 270, 365

[322] Miller T. B., van den Bosch F. C., Green S. B., Ogiya G., 2020, MNRAS,

[323] Million E. T., Allen S. W., 2009, MNRAS, 399, 1307

[324] Miyamoto M., Nagai R., 1975, PASJ, 27, 533

[325] Miyatake H., et al., 2019, ApJ, 875, 63

[326] Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319

[327] Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge University Press

[328] Mohr J. J., Fabricant D. G., Geller M. J., 1993, ApJ, 413, 492

[329] Molnar S. M., Hearn N., Haiman Z., Bryan G., Evrard A. E., Lake G., 2009, ApJ, 696, 1640

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

[331] Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5

[332] Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19

[333] Morandi A., Sun M., Forman W., Jones C., 2015, MNRAS, 450, 2261

[334] More S., van den Bosch F. C., Cacciato M., Skibba R., Mo H. J., Yang X., 2011, MNRAS, 410, 210

[335] More S., Diemer B., Kravtsov A. V., 2015, ApJ, 810, 36

[336] Morinaga Y., Ishiyama T., 2020, MNRAS, 495, 502

[337] Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

[338] Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822

[339] Motl P. M., Hallman E. J., Burns J. O., Norman M. L., 2005, ApJ, 623, L63

[340] Mroczkowski T., et al., 2019, Space Sci. Rev., 215, 17

[341] Muldrew S. I., Pearce F. R., Power C., 2011, MNRAS, 410, 2617

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

[343] Nadler E. O., Mao Y.-Y., Green G. M., Wechsler R. H., 2019, ApJ, 873, 34

[344] Nadler E. O., et al., 2020, ApJ, 893, 48

[345] Nadler E. O., et al., 2021, Phys. Rev. Lett., 126, 091101

[346] Nagai D., 2006, ApJ, 650, 538

[347] Nagai D., Kravtsov A. V., 2005, ApJ, 618, 557

[348] Nagai D., Lau E. T., 2011, ApJ, 731, L10

[349] Nagai D., Kravtsov A. V., Kosowsky A., 2003, ApJ, 587, 524

[350] Nagai D., Vikhlinin A., Kravtsov A. V., 2007, ApJ, 655, 98

[351] Nagarajan A., et al., 2019, MNRAS, 488, 1728

[352] Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

[353] Necib L., et al., 2020, Nature Astronomy, 4, 1078

[354] Nelson K., Rudd D. H., Shaw L., Nagai D., 2012, ApJ, 751, 121

[355] Nelson K., Lau E. T., Nagai D., Rudd D. H., Yu L., 2014a, ApJ, 782, 107

[356] Nelson K., Lau E. T., Nagai D., 2014b, ApJ, 792, 25

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

[358] Ngan W. H. W., Carlberg R. G., 2014, ApJ, 788, 181

[359] Nierenberg A. M., et al., 2017, MNRAS, 471, 2224

[360] Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, ApJ, 803, 50

[361] Ntampaka M., Trac H., Sutherland D. J., Fromenteau S., Póczos B., Schneider J., 2016, ApJ, 831, 135

[362] Ntampaka M., et al., 2019a, ApJ, 876, 82

[363] Ntampaka M., Rines K., Trac H., 2019b, ApJ, 880, 154

[364] Ogiya G., 2018, MNRAS, 480, L106

[365] Ogiya G., Burkert A., 2016, MNRAS, 457, 2164

[366] Ogiya G., Mori M., 2014, ApJ, 793, 46

[367] Ogiya G., Mori M., Miki Y., Boku T., Nakasato N., 2013, in Journal of Physics Conference Series. p. 012014, doi:10.1088/1742-6596/454/1/012014

[368] Ogiya G., van den Bosch F. C., Hahn O., Green S. B., Miller T. B., Burkert A., 2019, MNRAS, 485, 189

[369] Oguri M., Lee J., 2004, MNRAS, 355, 120

[370] Oguri M., Takada M., Umetsu K., Broadhurst T., 2005, ApJ, 632, 841

[371] Oh S.-H., de Blok W. J. G., Brinks E., Walter F., Kennicutt Jr. R. C., 2011, AJ, 141, 193

[372] Okabe N., Okura Y., Futamase T., 2010, ApJ, 713, 291

[373] Okabe N., Futamase T., Kajisawa M., Kuroshima R., 2014, ApJ, 784, 90

[374] Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920

[375] Oman K. A., et al., 2015, MNRAS, 452, 3650

[376] Onions J., et al., 2012, MNRAS, 423, 1200

[377] Osato K., Flender S., Nagai D., Shirasaki M., Yoshida N., 2018, MNRAS, 475, 532

[378] Osato K., Shirasaki M., Miyatake H., Nagai D., Yoshida N., Oguri M., Takahashi R., 2020, MNRAS, 492, 4780

[379] Ostriker J. P., Peebles P. J. E., 1973, ApJ, 186, 467

[380] Ostriker J. P., Spitzer Lyman J., Chevalier R. A., 1972, ApJ, 176, L51

[381] Ota N., Nagai D., Lau E. T., 2018, PASJ, 70, 51

[382] Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

[383] Parrish I. J., Stone J. M., Lemaster N., 2008, ApJ, 688, 905

[384] Parrish I. J., McCourt M., Quataert E., Sharma P., 2012, MNRAS, 419, L29

[385] Peñarrubia J., Benson A. J., 2005, MNRAS, 364, 977

[386] Peñarrubia J., Navarro J. F., McConnachie A. W., 2008, ApJ, 673, 226

[387] Peñarrubia J., Navarro J. F., McConnachie A. W., Martin N. F., 2009, ApJ, 698, 222

[388] Pedregosa F., et al., 2011, J. Mach. Learn. Res., 12, 2825

[389] Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton University Press

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

[391] Pieri L., Bertone G., Branchini E., 2008, MNRAS, 384, 1627

[392] Pike S. R., Kay S. T., Newton R. D. A., Thomas P. A., Jenkins A., 2014, MNRAS, 445, 1774

[393] Pillepich A., Porciani C., Reiprich T. H., 2012, MNRAS, 422, 44

[394] Pillepich A., et al., 2018, MNRAS, 473, 4077

[395] Pillepich A., et al., 2019, MNRAS, 490, 3196

[396] Plackett R. L., 1950, Biometrika, 37, 149

[397] Planck Collaboration et al., 2011, A&A, 536, A9

[398] Planck Collaboration et al., 2014a, A&A, 571, A16

[399] Planck Collaboration et al., 2014b, A&A, 571, A20

[400] Planck Collaboration et al., 2016a, A&A, 594, A13

[401] Planck Collaboration et al., 2016b, A&A, 594, A24

[402] Planck Collaboration et al., 2020, A&A, 641, A6

[403] Planelles S., Borgani S., Dolag K., Ettori S., Fabjan D., Murante G., Tornatore L., 2013, MNRAS, 431, 1487

[404] Planelles S., et al., 2017, MNRAS, 467, 3827

[405] Plummer H. C., 1911, MNRAS, 71, 460

[406] Pontzen A., Governato F., 2012, MNRAS, 421, 3464

[407] Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

[408] Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

[409] Pratt G. W., Croston J. H., Arnaud M., Böhringer H., 2009, A&A, 498, 361

[410] Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., 2019, Space Sci. Rev., 215, 25

[411] Press W. H., Schechter P., 1974, ApJ, 187, 425

[412] Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical recipes in FORTRAN. The art of scientific computing. Cambridge University Press

[413] Primack J. R., Gross M. A. K., 2001, Hot dark matter in cosmology. Springer-Verlag Berlin Heidelberg, pp 287–308

[414] Pullen A. R., Benson A. J., Moustakas L. A., 2014, ApJ, 792, 24

[415] Purcell C. W., Zentner A. R., 2012, J. Cosmology Astropart. Phys., 2012, 007

[416] Quinlan J. R., 1986, Mach. Learn., 1, 81

[417] Ragagnin A., Dolag K., Biffi V., Cadolle Bel M., Hammer N. J., Krukau A., Petkova M., Steinborn D., 2017, A&C, 20, 52

[418] Raghunathan S., et al., 2019, ApJ, 872, 170

[419] Rasia E., et al., 2006, MNRAS, 369, 2013

[420] Rasia E., Meneghetti M., Ettori S., 2013, AstRv, 8, 40

[421] Rasia E., et al., 2014, ApJ, 791, 96

[422] Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006a, MNRAS, 367, 387

[423] Read J. I., Goerdt T., Moore B., Pontzen A. P., Stadel J., Lake G., 2006b, MNRAS, 373, 1451

[424] Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30

[425] Relatores N. C., et al., 2019, ApJ, 887, 94

[426] Remus R.-S., Dolag K., Naab T., Burkert A., Hirschmann M., Hoffmann T. L., Johansson P. H., 2017, MNRAS, 464, 3742

[427] Retana-Montenegro E., van Hese E., Gentile G., Baes M., Frutos-Alfaro F., 2012, A&A, 540, A70

[428] Rico J., 2020, Galaxies, 8, 25

[429] Riess A. G., et al., 1998, AJ, 116, 1009

[430] Robles V. H., et al., 2017, MNRAS, 472, 2945

[431] Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, MNRAS, 430, 81

[432] Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651

[433] Salvati L., Douspis M., Ritz A., Aghanim N., Babul A., 2019, A&A, 626, A27

[434] Samuel J., et al., 2020, MNRAS, 491, 1471

[435] Santos J. S., Rosati P., Tozzi P., Böhringer H., Ettori S., Bignamini A., 2008, A&A, 483, 35

[436] Sawala T., et al., 2016, MNRAS, 457, 1931

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

[438] Sayers J., et al., 2019, ApJ, 880, 45

[439] Schellenberger G., Reiprich T. H., Lovisari L., Nevalainen J., David L., 2015, A&A, 575, A30

[440] Schneider A., Stoira N., Refregier A., Weiss A. J., Knabenhans M., Stadel J., Teyssier R., 2020, J. Cosmology Astropart. Phys., 2020, 019

[441] Schumann M., 2019, Journal of Physics G Nuclear Physics, 46, 103003

[442] Sehgal N., et al., 2019, in BAAS. p. 6 (arXiv:1906.10134)

[443] Sembolini F., Yepes G., De Petris M., Gottlöber S., Lamagna L., Comis B., 2013, MNRAS, 429, 323

[444] Sereno M., Ettori S., Moscardini L., 2015, MNRAS, 450, 3649

[445] Shaw L. D., Nagai D., Bhattacharya S., Lau E. T., 2010, ApJ, 725, 1452

[446] Shi X., Komatsu E., 2014, MNRAS, 442, 521

[447] Shi X., Zhang C., 2019, MNRAS, 487, 1072

[448] Shi X., Komatsu E., Nelson K., Nagai D., 2015, MNRAS, 448, 1020

[449] Shi X., Komatsu E., Nagai D., Lau E. T., 2016, MNRAS, 455, 2936

[450] Shi X., Nagai D., Lau E. T., 2018, MNRAS, 481, 1075

[451] Shi X., Nagai D., Aung H., Wetzel A., 2020, MNRAS, 495, 784

[452] Shirasaki M., Nagai D., Lau E. T., 2016, MNRAS, 460, 3913

[453] Shirasaki M., Macias O., Ando S., Horiuchi S., Yoshida N., 2020a, Phys. Rev. D, 101, 103022

[454] Shirasaki M., Lau E. T., Nagai D., 2020b, MNRAS, 491, 235

[455] Shu Y., et al., 2015, ApJ, 803, 71

[456] Sijacki D., Pfrommer C., Springel V., Enßlin T. A., 2008, MNRAS, 387, 1403

[457] Silk J., 1968, ApJ, 151, 459

[458] Simionescu A., et al., 2019a, arXiv e-prints, p. arXiv:1908.01778

[459] Simionescu A., et al., 2019b, Space Sci. Rev., 215, 24

[460] Singh P., Saro A., Costanzi M., Dolag K., 2020, MNRAS, 494, 3728

[461] Sofue Y., 2013, PASJ, 65, 118

[462] Somalwar J. J., Chang L. J., Mishra-Sharma S., Lisanti M., 2021, ApJ, 906, 57

[463] Somerville R. S., et al., 2018, MNRAS, 473, 2714

[464] Spergel D. N., et al., 2003, ApJS, 148, 175

[465] Spitzer Jr. L., 1958, ApJ, 127, 17

[466] Spitzer L., 1987, Dynamical evolution of globular clusters. Princeton University Press

[467] Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

[468] Springel V., et al., 2005, Nature, 435, 629

[469] Springel V., et al., 2008, MNRAS, 391, 1685

[470] Stanek R., Rasia E., Evrard A. E., Pearce F., Gazzola L., 2010, ApJ, 715, 1508

[471] Steinborn L. K., Dolag K., Hirschmann M., Prieto M. A., Remus R.-S., 2015, MNRAS, 448, 1504

[472] Steinborn L. K., Dolag K., Comerford J. M., Hirschmann M., Remus R.-S., Teklu A. F., 2016, MNRAS, 458, 1013

[473] Stref M., Lavalle J., 2017, Phys. Rev. D, 95, 063003

[474] Strigari L. E., Koushiappas S. M., Bullock J. S., Kaplinghat M., 2007, Phys. Rev. D, 75, 083526

[475] Sunayama T., More S., 2019, MNRAS, 490, 4945

[476] Sunyaev R. A., Zeldovich Y. B., 1972, CoASP, 4, 173

[477] Taffoni G., Mayer L., Colpi M., Governato F., 2003, MNRAS, 341, 434

[478] Tashiro M., et al., 2018, in Proc. SPIE. p. 1069922, doi:10.1117/12.2309455

[479] Taylor J. E., Babul A., 2001, ApJ, 559, 716

[480] Taylor J. E., Babul A., 2004, MNRAS, 348, 811

[481] Teklu A. F., Remus R.-S., Dolag K., Beck A. M., Burkert A., Schmidt A. S., Schulze F., Steinborn L. K., 2015, ApJ, 812, 29

[482] Tibshirani R., 1996, J. Royal Stat. Soc. B, 58, 267

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

[484] Tollet E., et al., 2016, MNRAS, 456, 3542

[485] Tollet É., Cattaneo A., Mamon G. A., Moutard T., van den Bosch F. C., 2017, MNRAS, 471, 4170

[486] Tonini C., Lapi A., Salucci P., 2006, ApJ, 649, 591

[487] Tormen G., 1997, MNRAS, 290, 411

[488] Tormen G., Bouchet F. R., White S. D. M., 1997, MNRAS, 286, 865

[489] Tormen G., Diaferio A., Syer D., 1998, MNRAS, 299, 728

[490] Trac H., Bode P., Ostriker J. P., 2011, ApJ, 727, 94

[491] Tremmel M., et al., 2019, MNRAS, 483, 3336

[492] Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16

[493] Vale A., Ostriker J. P., 2006, MNRAS, 371, 1173

[494] Van Waerbeke L., Hinshaw G., Murray N., 2014, Phys. Rev. D, 89, 023508

[495] Vattis K., Toomey M. W., Koushiappas S. M., 2020, arXiv e-prints, p. arXiv:2008.11577

[496] Vazza F., Brüggen M., Gheller C., 2013, MNRAS, 428, 2366

[497] Vazza F., Angelinelli M., Jones T. W., Eckert D., Brüggen M., Brunetti G., Gheller C., 2018, MNRAS, 481, L120

[498] Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, Nature, 481, 341

[499] Vegetti S., Koopmans L. V. E., Auger M. W., Treu T., Bolton A. S., 2014, MNRAS, 442, 2017

[500] Ventimiglia D. A., Voit G. M., Donahue M., Ameglio S., 2008, ApJ, 685, 118

[501] Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Phys. Rev. D, 71, 063534

[502] Vogelsberger M., Zavala J., Loeb A., 2012, MNRAS, 423, 3740

[503] Walker S., et al., 2019, Space Sci. Rev., 215, 7

[504] Wang H. Y., Jing Y. P., Mao S., Kang X., 2005, MNRAS, 364, 424

[505] Wang Y., Yang X., Mo H. J., Li C., van den Bosch F. C., Fan Z., Chen X., 2008, MNRAS, 385, 1511

[506] Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, MNRAS, 454, 83

[507] Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52

[508] Weinberg M. D., 1989, MNRAS, 239, 549

[509] Weinberg M. D., Katz N., 2007, MNRAS, 375, 425

[510] Wetzel A. R., 2011, MNRAS, 412, 49

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

[512] White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

[513] Widrow L. M., 2000, ApJS, 131, 39

[514] Wik D. R., Sarazin C. L., Ricker P. M., Rand all S. W., 2008, ApJ, 680, 17

[515] Wu H.-Y., Hahn O., Wechsler R. H., Behroozi P. S., Mao Y.-Y., 2013, ApJ, 767, 23

[516] Xhakaj E., Diemer B., Leauthaud A., Wasserman A., Huang S., Luo Y., Adhikari S., Singh S., 2020, MNRAS, 499, 3534

[517] Yang X., van den Bosch F. C., Mo H. J., Mao S., Kang X., Weinmann S. M., Guo Y., Jing Y. P., 2006, MNRAS, 369, 1293

[518] Yang S., Du X., Benson A. J., Pullen A. R., Peter A. H. G., 2020, MNRAS, 498, 3902

[519] Yu L., Nelson K., Nagai D., 2015, ApJ, 807, 12

[520] Zel'Dovich Y. B., 1970, A&A, 500, 13

[521] Zentner A. R., Bullock J. S., 2003, ApJ, 598, 49

[522] Zentner A. R., Berlind A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005a, ApJ, 624, 505

[523] Zentner A. R., Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2005b, ApJ, 629, 219

[524] Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

[525] Zhang J., Fakhouri O., Ma C.-P., 2008, MNRAS, 389, 1521

[526] Zhang Y.-Y., et al., 2010, ApJ, 711, 1033

[527] Zhang Y.-Y., Andernach H., Caretta C. A., Reiprich T. H., Böhringer H., Puchwein E., Sijacki D., Girardi M., 2011, A&A, 526, A105

[528] Zhang T., Liao S., Li M., Gao L., 2019, MNRAS, 487, 1227

[529] Zhao H., 1996, MNRAS, 278, 488

[530] Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, ApJ, 707, 354

[531] Zheng Z., et al., 2005, ApJ, 633, 791

[532] Zhuravleva I., Churazov E., Kravtsov A., Lau E. T., Nagai D., Sunyaev R., 2013, MNRAS, 428, 3274

[533] Zhuravleva I., et al., 2014, Nature, 515, 85

[534] Zinger E., Dekel A., Kravtsov A. V., Nagai D., 2018, MNRAS, 475, 3654

[535] Zolotov A., et al., 2012, ApJ, 761, 71

[536] Zu Y., Mandelbaum R., Simet M., Rozo E., Rykoff E. S., 2017, MNRAS, 470, 551

[537] Zubeldia Í., Challinor A., 2019, MNRAS, 489, 401

[538] Zwicky F., 1933, Helv. Phys. Acta, 6, 110

[539] da Silva A. C., Kay S. T., Liddle A. R., Thomas P. A., 2004, MNRAS, 348, 1401

[540] de Haan T., et al., 2016, ApJ, 832, 95

[541] van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015, ApJ, 798, L45

[542] van Kampen E., 1995, MNRAS, 273, 295

[543] van den Bosch F. C., 2002, MNRAS, 331, 98

[544] van den Bosch F. C., 2017, MNRAS, 468, 885

[545] van den Bosch F. C., Jiang F., 2016, MNRAS, 458, 2870

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

[547] van den Bosch F. C., Swaters R. A., 2001, MNRAS, 325, 1017

[548] van den Bosch F. C., Yang X., Mo H. J., Norberg P., 2005a, MNRAS, 356, 1233

[549] van den Bosch F. C., Tormen G., Giocoli C., 2005b, MNRAS, 359, 1029

[550] van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, MNRAS, 430, 725

[551] van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713

[552] van den Bosch F. C., Jiang F., Campbell D., Behroozi P., 2016, MNRAS, 455, 158

[553] van den Bosch F. C., Ogiya G., Hahn O., Burkert A., 2018, MNRAS, 474, 3043

[554] van den Bosch F. C., Lange J. U., Zentner A. R., 2019, MNRAS, 488, 4984

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

[556] von der Linden A., et al., 2014a, MNRAS, 439, 2

[557] von der Linden A., et al., 2014b, MNRAS, 443, 1973