Coupling Semi-Analytic and N-body Galaxies (CoSANG) for cosmological stellar halo simulations I- Methods and the structure of dark matter halos

Shahram Talei\textsuperscript{1},* Krista Mccord\textsuperscript{2}, Jeremy Bailin\textsuperscript{1}, Darren J. Croton\textsuperscript{3,4} Alexandra Mannings\textsuperscript{5}, Michael Sitarz\textsuperscript{6}, Annelia Anderson \textsuperscript{1} Brooke Bailey \textsuperscript{1}, Alexis Rollins \textsuperscript{1}

\textsuperscript{1}Department of Physics and Astronomy, University of Alabama Box 870324, Tuscaloosa, AL, 35487-0324, USA
\textsuperscript{2}Deep Space Systems Inc, 8100 Shaffer Pkwy, Littleton, CO 80127, USA
\textsuperscript{3}Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia
\textsuperscript{4}ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)
\textsuperscript{5}Department of Astronomy & Astrophysics, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
\textsuperscript{6}Department of Physics & Astronomy, University of Kansas, Lawrence, KS 66045, USA

Abstract

We present CoSANG (Coupling Semi-Analytic and N-body Galaxies), a new hybrid model for cosmological dark matter and stellar halo simulations. In this approach a collisionless model (Gadget3) is used for gravitational interactions while a coupled semi-analytic model (SAGE) calculates baryonic effects at each time-step. This live self-consistent interaction at each time-step is the key difference between CoSANG and traditional semi-analytic models that are mainly used for post-processing. By accounting for the gravitational effect of the baryons, CoSANG can overcome some of the deficiencies of pure N-body simulations, while being less computationally expensive than hydrodynamic simulations. Moreover CoSANG can produce stellar halo populations via tagging tracer dark matter particles. We demonstrate the performance and dynamical accuracy of this approach using both controlled test simulations and a set of three cosmological zoom-in simulations of Milky Way (MW) mass halos. We simulate each target halo both without the coupling (dark matter only, hereafter DMO) and with the coupling (CoSANG). We compare the internal structure (subhalo distribution, shape and orientation) of the halos. The following changes are observed in the CoSANG model compared to the DMO model: 1) the total number of subhaloes close to the center of the halo is reduced, 2) the $V_\text{max}$ distribution peaks at a lower value and lies below the DMO model, 3) the axis ratio is smaller. The difference between DMO and CoSANG simulations is more significant in early forming halos.

Keywords: software: simulations – galaxies: evolution – (cosmology:) dark matter – methods: numerical

1 INTRODUCTION

Over recent decades a combination of precise cosmological measurements and accurate theoretical models have revolutionized our understanding of galaxy formation. The $\Lambda$CDM paradigm has become the most popular model and has been developed into a successful framework that explains many different cosmological observations with a remarkable precision, from the cosmic microwave background (e.g. Bennett et al. 2013; Planck Collaboration et al. 2016), to galaxy formation (e.g. Eggen et al. 1962; White 1984; Cole et al. 2000; Benson 2010; Silk & Mamon 2012). Nevertheless many aspects of galaxy formation are still highly debated.

Hierarchical structure formation play an important role in the $\Lambda$CDM scenario and galaxies build up their dynamical mass via dynamical accretion and mergers. Since dark matter acts as a potential well in the assembly of galaxies (White & Rees 1978), the distribution of the substructures and shape of the dark matter halo reflect their formation history. Therefore understanding the inner
structure of the dark matter halo can help us to understand galaxy formation within our cosmological framework. Detection of these substructures can confirm key predictions of ΛCDM theory and rule out many alternative theories.

According to this bottom-up scenario, structures form out of primordial Gaussian fluctuations in the density distribution in the early universe and then grow to the current size. Although structures grow linearly in the early universe (Zel’Dovich 1970), growth quickly becomes nonlinear with no known analytic solution, requiring numerical techniques to study. Gravitational interactions between all components govern structure formation (Peebles 1974, 1982; Davis et al. 1985).

A simplified two stage model can capture many aspects of this interaction between baryons and the dark matter (White & Rees 1978, for a more recent review see Oser et al. 2010). According to this model, the majority of stars are formed while falling in the dark matter potential well and therefore the invisible dark matter leaves a fingerprint on the visible stellar population. The gravitational potential of the baryons can alter the dark matter distribution in return (e.g. Garrison-Kimmel et al. 2014; Bryan et al. 2013; Zhu et al. 2016; Graus et al. 2019; Kelley et al. 2019); the mutual interactions between these components is crucial.

Since dark matter does not interact via the electromagnetic force, any visible component that traces the dark matter can provide a valuable probe of this distribution and be a bridge between observation and formation history. Galaxies undergo different mechanisms like violent relaxation, phase mixing, chaotic mixing and Landau damping (Lynden-Bell 1967, Shu 1978, Tremaine et al. 1986, Binney & Tremaine 2008) and most of the formation history is lost in the main body of the galaxy. We can study galaxy formation either by looking at early galaxies at higher redshifts or we can extract the formation history through cosmological archaeology in the local universe at lower redshifts. Galactic components with a very long dynamical time can help with the later approach because unlike the other components, they are not relaxed.

Stellar halos are a clear example of hierarchical accretion. The stellar halo is the most extended stellar component of a galaxy and it is the remnant of disruption. Although the stellar halo is a small fraction of stars in a galaxy, it displays a rich multi-component inner structure. The stellar halo of the Milky Way (MW) contains very old and metal-poor stars. It has a much longer relaxation time than the disk, and plays a key role as a time capsule preserving the formation history of the galaxy. The stellar halo can interact with the dark matter halo directly, and therefore can help us to understand the dark matter and determine the stellar content of the galaxy (Eggen et al. 1962; Zaritsky & White 1994; Helmi & White 1999; Johnston et al. 2008; Kuhlen et al. 2009; Vera-Ciro & Helmi 2013; Bonaca et al. 2014; Dai et al. 2018; Bonaca & Hogg 2018; D’Souza & Bell 2018; Zaritsky et al. 2020; Nadler et al. 2021; see Helmi 2020 for a recent review). Therefore a combined study of the dark matter halo and the stellar halo can shine a light on galaxy formation and bridge the gap between theory and observation.

Numerical simulations are extensively used to study galaxy formation and they are powerful tools to examine various models. Today these models can include multiple physical processes to reproduce more realistic galaxies; however despite many improvements in technology and the implementation of new efficient algorithms, simulating galaxies and their stellar halos is still extremely computationally expensive. This regime is very sensitive to the numerical parameters we choose (Power et al. 2003) and limited by numerical resolution and systematic errors (Bailin et al. 2014).

Any computational modeling is an attempt to improve resolution, cosmological sample size and physics (the number of processes that are included). Increasing any of these factors is constrained by computational resources. Depending on the questions that one wants to answer, different combinations of the simulation set-up can be optimized. However, most of the time this optimum requires compromising at least one other aspect of the simulation that could be important to explain observations.

Reproducing the relevant physical processes with fidelity is often the highest priority. Although dark matter is well described with collisionless gravitational interactions (even though the true nature of the dark matter is still unknown) (Kuhlen et al. 2012), baryonic matter interacts via multiple channels. Several recent studies include these processes, with a very high computational cost (e.g. Sanderson et al. 2018; Mackereth et al. 2019; Yu et al. 2020).

Many dark matter only (DMO) studies ignore baryons to reduce the computational cost. They may attempt to recover the missing structures through post-processing with an empirical model, but the missing baryon-dark matter interaction has a significant dynamical impact on dark matter structure and eventually on the stellar population. Full physics hydrodynamic models are not immune from these systematic errors either. In full physics simulations a sub-grid model can capture baryon-baryon interactions at sub-resolution scales.

A new fast and more accurate model that can produce a large galaxy sample at very high resolution would be an extremely powerful tool to understand galaxy formation. In this series of papers, we present one such approach: Coupled Semi-Analytic and N-body Galaxies (CoSANG), which takes advantage of a fast N-body code at its core, coupled with a well calibrated semi-analytic model (SAM) of galaxy formation. This model has limits but it is well suited for stellar halo simulations. It includes dynamical coupling between dark and baryonic components, which can produce more realistic dynamics, and tags stellar properties onto tracer dark matter particles do produce stellar halo populations.

CoSANG is used in three major studies: dark matter kinematics, stellar tracers kinematics and stellar halo populations. In this work we focus on the dark matter kinematics with CoSANG by studying dark matter halos, focusing on their shape; the abundance and the distribution of substructures; stellar halos produced by CoSANG will be discussed in upcoming work.

This paper is structured as follows: We start with a general overview of dark matter in Section 2 then summarize the observational results and theoretical models and simulations in Section 3. In Section 4 we present CoSANG and different components of the model and how we test them. In Section 5 we present dynamical results with both the DMO model and CoSANG and compare substructures and internal structure of the dark matter halos. Particle tagging with CoSANG and a stellar halo analysis will be discussed in a separate paper.

2 DARK MATTER HALOS

The gravitational collapse of dark matter halos can produce a universal profile (Einasto 1965, Einasto & Haardt 1989, Navarro et al. 1996), but the internal structure can vary between halos as a result of different formation histories and baryonic content (White & Rees 1978). These differences might manifest in the distribution of subhaloes, or the shape and orientation of the halos.

Subhaloes around the main halo have been studied extensively. Substructures can form and survive in small clumps even in very inner regions of the galactic halo (Diemand et al. 2008). The mass distribution of subhaloes can be described by a power-law distri-
bution (e.g. Helmi et al. 2002; Springel et al. 2008) or Schechter function (Schechter 1976).

Although many early studies showed discrepancies between observations of the substructures and predictions from ΛCDM and hydro models (e.g., Klypin et al. 1999, Moore et al. 1999), many of these discrepancies are reduced or disappear when including more and better baryonic physics in simulations (e.g. Bullock et al. 2000; Brooks et al. 2013; Wetzel et al. 2016). On the other hand the population of the observed satellites is growing rapidly and observational incompleteness could be a major concern (Tollerud et al. 2008; Walsh et al. 2009; Hargis et al. 2014; Kim et al. 2018).

Substructures within MW mass galaxies have been studied with ELVIS (Exploring the Local Volume in Simulations) in models of both zoom-in Local Group analogues and isolated structures (Garrison-Kimmel et al. 2014). They find local Group-like pairs average almost twice as many companions and the velocity field is kinematically hotter and more complex on Mpc scales. They report no difference in the abundance or kinematics of substructure within the virial radii of isolated versus paired hosts. Extension of this study into hydrodynamic models show that energetic feedback from supernovae and subsequent tidal stripping in hydro models significantly reduces the DM mass in the central regions of luminous satellite galaxies (Brooks & Zolotov 2014).

Another important parameter of dark matter haloes that can be affected by baryonic physics is its ellipsoidal shape. The most common method to define the shape of a halo is using the second moment of the mass distribution (Katz 1991; Dubinski & Carlberg 1991; Warren et al. 1992; Bailin et al. 2005; Allgood et al. 2006; Kuhlen et al. 2007; Zemp et al. 2011). Dark matter halo shapes have been studied in many recent simulations. Vera-Ciro et al. (2011) used the Aquarius N-body simulation (Springel et al. 2008) to demonstrate how the shape of individual halos changes with time, evolving from a typically prolate configuration at early stages to a more triaxial/oblate geometry at the present day. Baryonic effects were explored in the Illustris simulation by comparing Illustris with the Illustris dark model, and show that condensation of baryons results in significantly rounder and more oblate halos, with the median minor-to-major axis ratio $c/a = 0.7$ (Chua et al. 2019). A similar study with 30 zoom-in MW mass galaxies in the Auriga simulations confirms the same conclusion, and also finds a strong alignment between the disk and halo shape (Prada et al. 2019; see also Bailin et al. 2005).

3 HALO SIMULATIONS

Today highly scalable algorithms start from a random initial condition consistent with the CMB distribution and they are capable of reproducing a large variety of observable features. For detailed review of different codes and their capabilities, see Benson (2010); Vogelsberger et al. (2020); Angulo & Hahn (2022). The linear phase is evolved analytically in order to start the simulation at lower redshift where the nonlinear phase begins (e.g. Hahn & Abel 2011; Knebe et al. 2009).

Simulations can generally be categorized as either collisionless (N-body) or full physics (hydrodynamic) models. We provide here a summary of the simulation techniques that have been used to date that inform how CoSANG is constructed.

3.1 N-body and hydro models

N-body models only include gravitational interactions and ignore dissipationary baryonic physics. In general N-body models determine forces by solving Poisson’s equation. Although N-body models might include baryonic components, such as stars, they do not take into account non-gravitational effects.

N-body models are not restricted by sub-grid physics; they are much faster than hydrodynamic models, thus for a specific amount of computational power, they can achieve a higher resolutions. N-body models have been used extensively to study the local group (e.g. Besla et al. 2010; Erkal et al. 2018; Garavito Camargo et al. 2020; Vasiliev et al. 2021; Naidu et al. 2021).

The main disadvantage of N-body models is the lack of baryonic processes, therefore they are suitable for dark matter simulations (e.g. Navarro et al. 1996, 1997; Moore et al. 1999; Gao et al. 2004; Springel et al. 2008; Klypin et al. 2016) but they fail modeling baryonic contributions (e.g. galactic potentials, different feedback mechanisms, dissipation).

N-body models lacking baryons do not predict the abundance of subhaloes accurately due to effects like adiabatic contraction, tidal disruption, and reionization, which shape the dark matter distribution in both the main halo and its subhaloes (Zhu et al. 2016). In a comparison between DMO and hydrodynamic versions of the Illustris simulations (Vogelsberger et al. 2014a; Nelson et al. 2015), Graus et al. 2018 found that not only is the overall number of subhaloes reduced in the hydrodynamic simulation, but their radial distribution is less centrally concentrated, which suggests that dynamical disruption of substructure from the central host galaxy can reduce substructure counts compared to the DMO simulation.

One may improve N-body models adding baryonic effects. N-body models can include baryons such as stellar components (no interactions other than gravity, e.g. Garavito-Camargo et al. 2019), and if those components are themselves accurate then the N-body simulation will faithfully reproduce the dynamics.

Hydrodynamic models include full physics. The higher computational cost is the main disadvantage of these models. Subsequently small scale sub-physics is not resolved and an alternative sub-grid prescription is used to include important sub-grid processes such as star formation and cooling (e.g., Springel & Hernquist 2003). Sub-grid models are calibrated with different constraints (e.g. Schechter 1976) and they are capable of reproducing more realistic galaxies (e.g. Vogelsberger et al. 2014b; Schaye et al. 2015; Wang et al. 2015; Grand et al. 2017; Hopkins et al. 2018). Sub-grid models can impose a big restriction on the predictability of the model (Munshi et al. 2019; Li et al. 2020).

Examining 30 galaxies in the Auriga simulation shows a diverse inner structure and shapes and a connection between population gradients and mass assembly histories: galaxies with few significant progenitors have more massive halos, possess large negative halo metallicity gradients, and steeper density profiles (Monachesi et al. 2016; Monachesi et al. 2019).

3.2 Semi-analytic and hybrid models

Beside explicit simulation, implicit empirical models, also known as semi-analytic models (SAMs), have been extensively used as an independent technique or in combination with an explicit model to understand the underlying physical processes. In general SAMs calculate the evolution of the baryonic component analytically while the evolution of the dark matter component is either imported from an N-body model or produced using Monte Carlo technique (e.g.
White & Rees 1978; Cole 1991; Cole et al. 2000; White & Frenk 1991; Baugh 2006; Croton et al. 2006, 2016), or take advantage of machine learning (e.g., Behroozi et al. 2019).

SAMs can produce remarkably consistent results at much lower computational cost compared to hydrodynamic models (Benson et al. 2001). The main advantage of using SAMs is that these models are not restricted by resolution. SAMs are complimentary models and they rely heavily on observational data or explicit models as their input and for calibration.

Any combination of SAMs and N-body models can belong to one of these categories:

- Passive SAMs: No interaction; SAMs are used independently or for post-processing without any live dynamical interactions (Cole et al. 2000)
- Active SAMs: One side interaction; In these models an analytic baryonic galaxy is included in the dynamics but this galaxy does not evolve according to the dark matter halo evolution even though there might be a prescription for the time evolution. This technique is more common in non-cosmological simulations, for example stellar streams kinematic, such as Sagittarius, LMC, and the Milky Way (e.g. see Vasiliev et al. 2021 and references there).

- Interactive SAMs: Mutual interaction; There is a mutual interaction between an explicit simulation and the SAM. There is not many example in this category and perhaps the closest examples are Garrison-Kimmel et al. (2017) and Kelley et al. (2019), however, in those cases the growth of the galaxy is calculated using halo abundance (Behroozi et al. 2010) rather than a SAM.

Empirical models can open a new window into low cost stellar halo modeling, when they are coupled with an N-body model using a technique known as particle tagging (Bullock & Johnston 2005; Le Bret et al. 2017). Justified by the similarities between accreted dark matter particles and stellar halo populations, a sample of the virialized particles are modified and tagged as stars in a non-cosmological DMO simulation (Bullock & Johnston 2005). This model has been successfully used, for example in recovering accretion histories with substructure abundance (Johnston et al. 2008) and chemical space abundance (Font et al. 2006) using a prescription to correlate accretion history and enrichment due to different enrichment mechanisms (Robertson et al. 2005). This prescription is extended into cosmological simulations (Cooper et al. 2010; Rashkov et al. 2012; Cooper et al. 2017).

Comparison between observations and hydrodynamic simulations confirms that accretion-only models produce promising results at much lower costs. Application of particle tagging in both hydrodynamic models and N-body models, and side by side comparison with the stellar halo population directly produced by a hydrodynamic model, shows the existence of systematic errors (Bailin et al. 2014) although it is possible to reduce the error by improving the SAM and tagging scheme (Cooper et al. 2017).

These accretion based approaches have three assumptions in common (Le Bret et al. 2017):

- Stars form tightly bound to their parent halos.
- Recently formed star particles and their tagged DM analogues subsequently follow similar phase space trajectories.
- Baryonic effects are not important in shaping the stellar halo.

The third assumption is more problematic. Direct comparison between hydrodynamic models and N-body models confirms the significance of these baryonic effects Brooks & Zolotov 2014; Wetzel et al. 2016; Zhu et al. 2016; Samuel et al. 2020. We can test the effect of a baryonic disk by adding a galactic potential into a DMO model and compare the results with the same model with no disk. This potential make itself apparent in substructure distributions and the inner structure and the shape of the halo. Nearby satellite halos within the virial radius of the host halo have lower survival rates in the presence of a disk (D’Onghia et al. 2010; Bauer et al. 2018; Graus et al. 2019; Kelley et al. 2019) and this baryonic potential combined with different feedback mechanisms could possibly explain the discrepancy between the number of dwarf satellites in DMO and hydrodynamic simulations.

A hybrid scheme can significantly improve the accuracy and the range of problems we can study with fast N-body models. Adding an evolving analytic galactic potential into a collisionless model using abundance matching (Behroozi et al. 2013) results in significantly less substructures going back ≈ 8 Gyr that is a better fit to pericentres (minimum separation between the host and the subhalo) calculations from Gaia observations and demonstrate an opposite missing satellite problem (Kim et al. 2018; Kelley et al. 2019), rather than too many subhaloes within 50 kpc of the MW that had $V_{peak} \geq 20$ km/s to account for the number of ultra-faint galaxies already known within that volume today (Graus et al. 2019). There have been recent attempts towards hybrid stellar halo modeling (e.g. Ricotti et al. 2022), but CoSANG is the first self-consistent hybrid model for stellar halo simulations with a full baryon prescription consistent with dark matter evolution.

4 COSANG, STRUCTURE AND METHODS

CoSANG is a computationally inexpensive algorithm that takes advantage of a fast collisionless model for the explicit component of the simulation.

There are two major factors that contribute to systematic errors in accretion-only models (Bailin et al. 2014):

- (i) Inaccurate dynamics: Ignoring baryonic potentials (e.g. bulge, disk, supermassive black hole) will result in a less accurate dynamical evolution (The importance of baryonic potentials is extensively discussed in Section 5).
- (ii) Particle tagging scheme: The spatial and kinematic distribution and properties of stellar tracer particles will have an impact on the predicted halo structure (Naidu et al. 2021).

CoSANG is an attempt to overcome these deficiencies. CoSANG is an interactive hybrid model. The two phase development of CoSANG addresses these errors.

- (i) Phase 1: Self-consistent dynamics: CoSANG uses a live interaction between the SAM and the N-body model in each time-step in a live simulation (Figure 1).
- (ii) Phase 2: Particle tagging: CoSANG implements a comprehensive SAM (Section 4.4) and a flexible tagging scheme (Talei et al. in preparation).

CoSANG merges several different pre-existing code components so that they work together as one cohesive whole. CoSANG includes: Gadget3 (see Springel 2005; Springel et al. 2020 for descriptions of v2 and v4 respectively) as the N-body code, SUBFIND (Springel et al. 2001) for the halo finder, L-Basetree and L-Halotree as the merger tree code, and the semi-analytic model SAGE (Croton et al. 2006, Croton et al. 2016). This allows the dark matter results from the N-body simulation to influence the results in the SAM. Importantly the SAM results are fed back into the N-body simulation as it runs to perturb the dark matter particles. This feedback loop
Figure 1. Anatomy of CoSANG. A CoSANG simulation consists of two phases: The live phase (grey shaded) and post-processing (white). In each time-step the N-body code Gadget calculates gravitational interactions. The halo finder Subfind extracts dark matter halo information, while Halotree extracts the merger tree and SAGE calculates the galaxy population using halos and the merger tree. An analytic potential is applied to affected particles and this happens in all time-steps. CoSANG produces the stellar halo in the second phase.

is continuous throughout a CoSANG simulation allowing both the dark matter particles and baryonic matter to evolve together.

In each time-step Gadget3 (Figure 1) calculates the gravitational force. Subfind is implemented in Gadget3 and provides halo information. At each snapshot after the first, which are output approximately every 50 million years, L-Halotree derives the merger tree. The SAM, SAGE, determines the galactic properties of each halo and is used to calculate an analytic gravitational force of the bulge, the black hole and the disk for all halos and subhaloes. We add this new term to the total force. This cyclic interaction continues at each time-step.

We discuss each code component in the next few sections. Beside the main algorithm, CoSANG is equipped with many new analysis codes that are developed and used for the first time and are publicly available

4.1 N-body model Gadget

The objective of N-body codes is to solve Poisson’s equation to compute the forces from the particle positions, advance the particles, and update their positions. Gadget is a well known parallel N-body TreePM code that uses MPI (message passing interface), which combines the gravity tree algorithm for calculating short-range forces with the particle mesh (PM) method for calculating long-range forces.

There are three important characteristics of the TreePM code: the grouping algorithm, the multipole expansion, and the opening criterion (Springel 2005). The grouping method used is an oct tree (Barnes & Hut 1986), which starts with a cube containing the entire mass distribution. It then breaks up the main cube into smaller and smaller cubes until each cube holds only one particle. Gadget uses this tree to incorporate gravitational forces using a hierarchical multipole expansion that groups the particles spatially such that the gravitational force on the particles is computed with $O(N \log N)$ interactions.

The potential calculation is split into two parts, one for particles at short range and the other for more distant particles (Ewald 1921; Hernquist et al. 1991). A spline kernel (Monaghan & Lattanzio 1985) is implemented so we have a Newtonian potential at long range and a modified Newtonian potential at short range. At short range, softening becomes important since the potential cannot be evaluated with finite timesteps when the acceleration diverges, so Newtonian potential of a point mass at zero separation becomes $-\frac{G M}{\epsilon}$ where $\epsilon$ is the force softening. The softening for a halo with a particular mass is represented by equation

$$\epsilon = 4 \frac{R_{vir}}{\sqrt{m_{\text{particle}}} \ M_{\odot}}$$  (1)
Gadget3 is not public but a new version (version 4) was published recently with new improvements such as the multi-range force calculation (Springel et al. 2020). CoSANG was developed before Gadget4 was published and there are fundamental changes in the new version (such as using C++ instead of C and object oriented structure instead of functional structure), which are inconsistent with other components we use.

4.2 Halo finder SUBFIND and Rockstar

Dark matter halo properties are retrieved by halo finders. The halo finder SUBFIND was chosen since it is conveniently merged with Gadget3. SUBFIND (Springel et al. 2001) combines two methods to find dark matter halos and their subhaloes. The first method is Friends-of-Friends (FOF) which links nearby particles together to locate large groups of dark matter particles (Davis et al. 1985) and then local overdensities are used to identify the substructures. Once SUBFIND locates all of the subhaloes within a parent halo, it uses an unbinding method to eliminate particles that are not bound to the halo. Particles with total positive (binding + kinetic) energy are iteratively removed from the particle list (Springel et al. 2001). After the unbinding process, halo properties such as velocity, virial mass, virial radius and angular momentum are calculated for the SUBFIND halos.

We use the Rockstar halo finder (Behroozi et al. 2012) in conjunction with SUBFIND for post-processing. Rockstar is a modern adaptive hierarchical refinement of friends-of-friends groups in 6D phase-space that avoids a few known issues with SUBFIND.

4.3 Merger trees

The merger tree code used in CoSANG is comprised of two components, L-Basertree and L-Halotree. These codes were created for constructing merger trees from SUBFIND and therefore readily accept the halo finder’s output. L-Basertree tracks halo particles between two snapshots to determine a halo’s descendant (Springel et al. 2005). A descendant halo can have multiple progenitors, but a progenitor can only be linked to one descendant. Particles that survive in a descendant halo are weighted based on their binding energy within the halo (Springel et al. 2005). The group of particles that have the highest binding energy are labeled as the descendant halo (Srisawat et al. 2013). This guards against halos that disappear in a snapshot but reappear in the next due to numerical effects by checking for halos across three consecutive snapshots (Springel et al. 2005; Srisawat et al. 2013). L-Halotree takes output from L-Basertree that contains the descendants for each snapshot and links them together to construct a complete merger history of the halos across the simulation. The merger tree can then be used by a SAM to determine when a galaxy forms and how it should evolve based on the halo merger history.

4.4 Semi-analytic model (SAM) SAGE

SAMS are developed as companions to direct simulations and they are widely used in different scenarios (Baugh 2006). We use SAM to refer to any non-explicit calculation in the simulation. SAGE (Semi-Analytic Galaxy Evolution) is the SAM code used by CoSANG to determine when and where galaxies should form during a simulation. SAGE uses a galaxy formation model that incorporates gas infall, AGN heating, gas cooling, ejected gas, satellite galaxies, and galaxy mergers (Croton et al. 2006, 2016).

Gas infall into each dark matter halo is based on the fraction of baryonic matter that is in the dark matter halo. As a dark matter halo forms, the baryons tend to fall into the halo due to their gravitational interaction. SAGE determines and tracks the growth of baryons for the halo. It does this by using the cosmic baryonic fraction to determine how many baryons should be present in the halo \( f_b \Delta M_{\text{vir}} \) and compares it to the previous timestep to track the growth (Croton et al. 2016). As a way to conserve the baryonic fractions during mergers, which can lead to artificial quenching, no more infalling baryons are allowed if the hot halo mass exceeds the cosmic fraction. The updated version has also more sophisticated prescription to avoid instant stripping.

SAGE follows the hot halo model where the halo gas has an isothermal density profile at the virial temperature, 

\[
T_{\text{vir}} = 39.5 \left( \frac{V_{\text{vir}}}{\text{km s}^{-1}} \right)^2 \text{K},
\]

where \( V_{\text{vir}} \) is the virial velocity of the halo (Croton et al. 2006, 2016). The gas then cools and condenses down to the center of the halo. The cooling rate for this process is calculated using the cooling time,

\[
t_{\text{cool}} = \frac{3}{2} \frac{\mu m_p kT}{\rho_g(r) \Lambda(T, Z)},
\]

where \( \rho \) is the gas density, \( \mu m_p \) is the mean mass and \( \Lambda(T, Z) \) is the cooling function. The cooling rate and its relation to the halo virial radius, determines if the infalling gas is in the cold accretion regime or the hot halo regime.

Stellar mass has dynamical impacts in CoSANG, and so the details of how it is determined are given below. The use of the stellar populations for particle tagging will be discussed in a future paper.

In CoSANG, SAGE has been incorporated as an interactive part of the simulation that receives parameters from the N-body code and calculates baryonic features and returns this information into the next step of the simulation. The main assumption is that as each dark matter halo collapses, its own baryonic counterparts collapse with it and the mass fraction in baryons associated with every halo is taken to be \( f_b = 0.17 \) (Bennett et al. 2013).

The required information per halo is:

(i) \( M_{\text{vir}} \), the halo virial mass and the number of particles in the halo or subhalo;
(ii) \( V_{\text{vir}} \) and \( R_{\text{vir}} \), virial velocity and virial radius;
(iii) \( V_{\text{max}} \), the maximum circular velocity of the halo;
(iv) The Cartesian position, velocity and spin vector of each halo.

The baryon content in each halo grows by \( f_b M_{\text{vir}} - m_b \) where \( m_b \) is the total mass of baryons present in the previous time-step.

Star formation is implemented assuming that only cold gas above a critical surface density threshold can form stars, where:

\[
m_{\text{crit}} = 3.8 \times 10^9 \left( \frac{V_{\text{vir}}}{200 \text{ km s}^{-1}} \right) \left( \frac{r_{\text{disk}}}{10 \text{kpc}} \right) M_\odot,
\]

and

\[
r_{\text{disk}} = \frac{3}{V_{\text{vir}}} \Lambda R_{\text{vir}}.
\]

This star formation prescription is based on the work of Kennicutt 1998 (see also Springel et al. 2001; De Lucia et al. 2004; Croton et al. 2006) and proceeds at a rate

\[
m_* = \alpha_S \left( \frac{m_{\text{cold}} - m_{\text{crit}}}{t_{\text{dyn, disk}}} \right).
\]
where $\alpha_{\text{SF}}$ is the star formation efficiency, $m_{\text{cold}}$ is the total mass of cold gas, and $t_{\text{dyn, disk}} = r_{\text{disk}}/v_{\text{vir}}$ is the disk dynamical time.

AGN feedback in SAGE (Croton et al. 2006) consists of two modes, a quasar mode and a radio mode. The purpose of the quasar mode is to grow the black hole at the center of the galaxy. Supermassive black holes grow during galaxy mergers both by merging with each other and by accretion of cold disc gas. The updated SAM in Croton et al. 2016 adds in feedback from quasar winds by the energy equation:

$$E_{\text{BH}, Q} = \frac{1}{2} \Delta M_{\text{BH}} c^2,$$

where $\Delta M_{\text{BH}}$ is the change in the black hole mass, $\kappa$ is the wind efficiency parameter, $\eta = 0.1$ is the standard efficiency with which inertial mass is liberated upon approaching the event horizon, and $c$ the speed of light.

The low-energy radio mode is the result of hot gas accretion on to a central supermassive black hole once a static hot halo has formed around the host galaxy of the black hole. This accretion is described by

$$\dot{m}_{\text{BH}, R} = \kappa_{\text{AGN}} \left( \frac{m_{\text{BH}}}{10^8 M_\odot} \right) \left( \frac{f_{\text{hot}}}{0.1} \right) \left( \frac{V_{\text{vir}}}{200 \text{Km s}^{-1}} \right)^3,$$

where $m_{\text{BH}}$ is the black hole mass, $f_{\text{hot}}$ is the fraction of the total halo mass in the form of hot gas, $V_{\text{vir}} \approx T_{\text{vir}}^{1/2}$ is the virial velocity of the halo, and $\kappa_{\text{AGN}}$ is the efficiency of accretion (with units of $M_\odot yr^{-1}$).

This feedback affects the evolution of the baryonic components, and therefore has dynamical consequences, but there is no direct feedback on individual particles in CoSANG.

Galaxy mergers play a major role in the evolution of galaxies in terms of their overall growth and morphology. SAGE treats mergers as either a minor merger or major merger, depending on the mass ratio ($f_{\text{merge}} = 0.3$). If the merging galaxy's mass is below a certain percentage of the main galaxy's mass, then it is considered to be a minor merger. In this case, the satellite galaxy's stellar mass is incorporated into the bulge mass and its cold gas mass is added to the disk of the main galaxy (Croton et al. 2006). In the case of a major merger, where both galaxy masses are similar to each other, the disk can be destroyed to form an elliptical galaxy and a major starburst begins.

Croton et al. (2016) have tested this model to be in good agreement with observations and hydrodynamic simulations. SAGE has been used for studies such as galactic morphology, galaxy luminosity function, stellar mass-gas metallicity relationship, black hole-bulge mass relation and was recently upgraded with dust modeling (Triani et al. 2020).

### 4.5 Galactic potentials

The gravitational effects of baryonic galaxies are taken into account by including additional forces on particles within the N-body simulation. We include forces due to the bulge, the black hole and the galactic disk.

We adapt the same force softening length as Gadget. For separation we include the softening length $\epsilon$ and we use $r_\epsilon$ for $|\vec{r}|$

where:

$$r_\epsilon = \sqrt{|\vec{r}|^2 + \epsilon^2}.$$  

Bulge: The bulge is represented by a Hernquist potential (Hernquist 1990) which is spherically symmetric. The potential follows the equation

$$\Phi_{\text{bulge}} = -\frac{GM_{\text{bulge}}}{|r| + a_{\text{bulge}}},$$

where $a_{\text{bulge}}$ is the scale radius, $M_{\text{bulge}}$ is the bulge mass and $\vec{r}$ is the vector from the galaxy center to the dark matter particle that the force is being calculated for. The bulge scale radius is assumed to follow a constant fraction of the disk scale length that SAGE derives (Carollo 2004). The bulge-to-disk scale length ratio has been observed to fall within the range of 0.2 to 0.02 (Carollo 1999). For CoSANG, an intermediate value of 0.1 is used (Carollo 2004).

Including the softening, this force becomes

$$\tilde{F}_{\text{bulge}} = -\frac{GM_{\text{bulge}}\vec{r}}{r_\epsilon (r_\epsilon + a_{\text{bulge}})}. $$

Supermassive black hole: The Plummer model is used to model the central black hole as a softened point source.

$$\tilde{F}_{\text{BH}} = -\frac{GM_{\text{BH}}\vec{r}}{r_\epsilon^3 (r_\epsilon + a_{\text{bh}})}, $$

where $M_{\text{BH}}$ is the black hole mass and $r_\epsilon$ is the softening included separation.

Galactic disk: Although galactic disks generally have exponential radial profiles (e.g. Courteau 1996; de Jong 1996; MacArthur et al. 2003; Pohlen & Trujillo 2006), there is no analytic solution for the gravitational force from an exponential disk. A good approximation is the sum of three Miyamoto-Nagai (Miyamoto & Nagai 1975; Smith et al. 2015) potentials. Each Miyamoto-Nagai potential has the following form:

$$\Phi(R, z) = -\frac{GM_{\text{MN}}}{\sqrt{R^2 + (a + \sqrt{b^2 + z^2})^2}},$$

where $M_{\text{MN}}$ is the total disc mass, $a$ is the radial scalelength, and $b$ is the vertical scaleheight, and the parameters of the three MN potentials as a function of disc scalelength and scaleheight are as described in Smith et al. (2015).

For every galaxy disk mass $M_{\text{d}}$, disk scale length $r_{\text{disk}}$, and disk scale height $h_{\text{disk}} = r_{\text{disk}}/9$ are taken from SAGE and used as input to calculate other parameters in Table 1 of Smith et al. (2015).

Multiple stability test have been performed to ensure that orbits are stable under the influence of this force both in isolation and inside CoSANG (see section 4.9.2).

Galaxies in CoSANG are assumed to be oriented perpendicular to the angular momentum or spin of the dark matter halo. Rotation matrices are applied to the particles to align the angular momentum vector to lie along the $z$ axis. The galaxy forces are then calculated on the particles within the halo and the forces are rotated back to the original frame.

### 4.6 Mass adjustment

CoSANG, as an N-body simulation, contains particles that initially represent both the dark matter and baryonic matter. By adding in a galactic potential based on the SAM, a fraction of mass needs to be subtracted from the dark matter particles to account for these baryons. Only particles that are within the halo virial radius have a percentage of mass subtracted for baryons. This percentage is based on the fraction of mass that an individual particle accounts for in the halo (Equation 14). This fraction is multiplied by the galaxy mass...
\[ f_m = \frac{m_{\text{particle}}}{M_{\text{vir}}} \]  

(14)

The total mass incorporates both the dark matter and baryonic matter. CoSANG distinguishes between the two by subtracting a fraction of mass from the dark matter particles. The mass subtracted from the dark matter particles represents the baryonic matter due to the galactic potential,

\[ \Delta m = f_m M_{\text{galaxy}}, \]  

(15)

where \( M_{\text{galaxy}} = M_{\text{BH}} + M_{\text{cold gas}} + M_{\text{stellar}} \) and where \( M_{\text{cold gas}} \) is the cold gas mass and \( M_{\text{stellar}} \) is the galaxy stellar mass from SAGE. Finally,

\[ m_{\text{particle}} \rightarrow m_{\text{particle}} - \Delta m. \]  

(16)

A particle can reside in both the main halo with a central galaxy and a halo with a satellite galaxy. Our method accounts for this case since the adjusted particle mass is preserved over the time-step. If a particle’s mass was adjusted once for the main halo, when it is adjusted again for a satellite, the adjusted mass is used in equation 15 and 16. After each time-step, the particle masses are also reset with the initial masses from the ICs. The masses are reset before the halo finder SUBFIND runs. The halo finder needs to run on the original particle masses since it needs to account for the total mass in the halo, otherwise the halo parameters will be inaccurate.

### 4.7 Time resolution and extrapolation

SAGE is run at each snapshot, which is when the galaxies and their parameters are updated. Snapshots are produced on average 50 million years apart. Between each snapshot, there are multiple time-steps when the forces are calculated for each dark matter particle. The galaxy properties should continue to evolve between snapshots, but until the next snapshot, we do not know what values of those properties should be. It is, however, important to take this evolution into account because large non-adiabatic jumps in the galaxy parameters can impact the halo unrealistically, therefore a smooth transition between snapshots is required. One way to do this is to extrapolate the next point for the galaxy parameter. Based on how the galaxies behave during the simulation up to a certain point we can predict how they should continue to evolve. A linear extrapolation is carried out on the SAGE galaxies from the previous two snapshots to determine how they should evolve to the next snapshot. The parameters extrapolated by CoSANG are Cold Gas Mass, Stellar Mass, Bulge Mass, Black Hole Mass, Disk Scale Radius and Angular Momentum Vector.

SAGE is run independently at each snapshot and has no connection to a previous SAGE run in terms of linking galaxies across snapshots. CoSANG therefore must determine which previous SAGE galaxy to use to define its extrapolation. CoSANG stores in memory the old galaxy data from the previous snapshot and the galaxy data from the current snapshot, and searches through the galaxy data from the two most recent snapshots to match galaxies within 60 kpc of each other.

The galaxy positions are based on the iterative centers of mass (Section 4.8), which are updated at each time-step and so they do not need to be extrapolated. Once galaxy matches have been found, all of the other properties are linearly extrapolated. The slope is calculated for each galaxy between the current snapshot and the previous one and is recalculated at each following snapshot. Galaxy properties are not extrapolated if the galaxy is new, no galaxy match is found or if the galaxy mass increases by more than 20% indicating a major merger occurred. Extrapolation for major mergers can cause an over extrapolation of the galaxy parameters. Major mergers are defined as having the mass ratio of 0.3 and above. SUBFIND, like many halo finder, fundamentally has problems finding host halos, especially during major mergers. This problem can result in mass fluctuations and ambiguity in distinguishing the host halo from merging subhaloes (Behroozi et al. 2013).

#### 4.8 Center of mass

In order for the baryonic potential to be stable, it must reside at the deepest point of the potential regardless of how the dark matter particles sample the halo structure. Tests with isolated spherically-symmetric halos reveal that placing the baryonic potential at the naive center of mass, or at the most bound particle, is not stable (McCord 2017).

In order to calculate an accurate center of mass for the halo, we use an iterative process. The algorithm searches for any particles that are within the virial radius of the halo and calculates the initial center of mass position. At each iteration, the sphere is shrunk by 20%, and a new center of mass for the particles within that sphere is calculated. The iteration continues until the difference between two consecutive iterations is less than a 1% threshold. Calculating the center of mass for all haloes at each time step adds significant computational expense, but is necessary for numerical stability.

The placement of galaxies at the center of mass of their host halos is a good approximation for most of the evolution but this assumption may not be very accurate during merger events, especially during major mergers in which dark matter and baryons may have different centers of mass.

Merger events pose a real challenge to N-body models with a FOF halo finder. These algorithms have difficulties distinguishing halos at close proximity. Moreover SAMs have difficulties extrapolating parameters during mergers in a live simulation and they may consider a close encounter as a merger and vice versa. SAMs can correct themselves in the next time-step and this ambiguity has significant dynamical effect but any merger events should be treated with more caution.

#### 4.9 Tests and performance

To ensure that CoSANG performs as expected, we have performed a number of tests, consisting of:

- Technical tests (performance, scalability, compatibility, efficiency)
- Scientific tests (dynamics, numerical errors)

Technical tests consist of using different hardware configurations, memory load balance, and scaling. Scientific tests include integrating the orbit of a test particle, stability under small perturbations, evolution of an isolated NFW halo, and performing a small box cosmological simulation at different mass resolution. We present one example of the most important tests in this section: the scaling test and an isolated NFW halo. See also the test particle orbital integration in Appendix A.
Figure 2. Performance of CoSANG as a function of mass resolution and number of compute cores. We run a zoom-in simulation of a target halo at different mass resolutions (red $10^6 M_\odot$, blue $10^5 M_\odot$, green $10^4 M_\odot$) and record the required time to accomplish specific tasks between two redshifts. Linear (perfect) scaling is shown in gray for comparison. Gadget simulations are shown as solid lines and CoSANG is shown with dashed lines. CoSANG is slower than Gadget by 50% or less. The highest resolution simulations cannot successfully run in the compute queue for lower numbers of cores.

For further details on all tests that have been performed on CoSANG, see McCord 2017.  

4.9.1 Scalability test

CoSANG is built on Gadget, which scales well as both the particle number and number of compute cores increase (Springel 2005). In order to determine the efficiency of the other parts of the code, such as the halo finder, SAM, center of mass calculation, and analytic gravitational force, we have performed a number of timing tests.

We test the scalability of the model by running the same model with Gadget (DMO) and CoSANG, using different number of processors and record the time they accomplish specific tasks. For ideal scaling, there would be a linear relationship between the number of processors and performance of the code, doubling the number of processors halving the runtime; however, in real world performance is always worse than ideal.

The most time consuming new piece of the algorithm is the iterative calculation of the center of mass which technically scales as $O(N)$. This test is done with a zoom-in box around one MW mass halo, starting at redshift 10 and continuing for less than 2 Gyr. This was enough to compare different models and running a full simulation for a longer time would have exceeded our benchmark allocation.

The required time for lower redshifts is much longer and it slows down more in CoSANG because there are more resolved structures and full grown halos contain very high numbers of particles.

We perform scaling tests using different number of processors and different resolutions. Results are shown in Figure 2. This plot shows the compute time versus the number of processors. We tested three mass resolutions, $10^4-6 M_\odot$ (different colors). Using small number of processors for the highest mass resolution it was not possible to complete the task within the supercomputer queue time limit; therefore, there is no point for finest mass resolution for fewer number of processors. No data is shown for $10^5-6 M_\odot$ at high number of cores because simulations run very fast and the recorded time is mostly the hardware input/output rate not numerical calculations. CoSANG requires 50%-100% more memory than Gadget due to using new variable and iterative calculations. CoSANG scales well up to ~ 500 cores for a mass resolution of $10^5 M_\odot$; further improvement in scaling is the subject of ongoing work.

4.9.2 Stability test

A critical test of CoSANG is that a disk potential embedded in an isolated dark matter halo will remain stable, with particles orbiting correctly in the combined dark matter + disk potential. (Binney & Tremaine 2008) and does not create numerical artifacts or non-physical orbits. The orbit of a single particle is discussed in Appendix A.

We generate an isolated NFW (Navarro et al. 1996) halo IC with Galstep 3 (e.g. Ruggiero 2017). This simulation is a non-cosmological model using CoSANG. We run the simulation using CoSANG and let particles evolve under the influence of the analytic potential and look for evolution in the density profile. We also run a cosmological MW mass halo, both with DMO and CoSANG and compare the density profile. The structure of the cosmological halo remains stable during entire simulation from redshift 20 to redshift 0.

5 RESULTS AND DISCUSSION

CoSANG is designed for stellar halo studies, and is therefore aimed at MW mass galaxies. We select a sample of initial conditions that will produce galaxies in the same mass range, but possibly with different histories. Our original complete sample covers an extended mass resolution $10^4-7 M_\odot$. Here we present the highest mass resolution while lower resolutions were used for convergence tests and calibrations. We focus on the dynamical effects of the baryonic potential on the dark matter substructures.

In this work we present three MW mass halos. We use zoom ICs from the FIRE project (Hopkins et al. 2018) that are known to produce MW mass galaxies in hydro simulations.

We reproduce the initial condition using MUSIC (Hahn & Abel 2011). The parameters are identical to those used by FIRE 4 (Wetzel et al. 2022), except that we use a starting redshift of 20, and do not create baryonic particles.

We run each simulation twice, one full CoSANG run and one where we turn off all CoSANG features (this is effectively Gadget3) as the DMO model. This allows us to understand the effect of the galactic potential on the dark matter distributions.

There are slight differences in the calculated parameters between SUBFIND and Rockstar (e.g. virial mass) but no significant differences between the two methods.

---

2 Test examples can be found at the following github repository: https://github.com/stalei

3 https://github.com/ruggiero/galstep

4 http://www.tapir.caltech.edu/~phopkins/publicICs/
5.1 Target halos and initial conditions

The cosmological zoom-in technique (Katz et al. 1994, Navarro & White 1994, Hahn & Abel 2011) is used for efficiency. In summary, the zoom-in technique starts with a uniform resolution box at low resolution. Target halos are selected and their particles are tracked back in the initial condition file. We then create a new set of initial condition with high resolution for the region surrounding these particles, leaving all particles outside this region at lower resolution.

Our cosmological parameters are as follows: $\Omega_M = 0.272$, $\Omega_{\Lambda} = 0.728$, $\Omega_{\text{baryon}} = 0.0455$, $h_0 = 0.702$ and a periodic box of size $60 h^{-1}$Mpc. The highest mass resolution has particles of mass $2.37266 \times 10^{5} M_\odot$ and force softening length $2$ kpc, starting at redshift 20 and running down to redshift 0. We extract halos at redshift 0. Our target MW mass halos have $M_{\text{halo}} \sim 10^{12} M_\odot$.

Particles outside the Lagrange volume have coarser resolution. Contamination and different biases can affect the zoom-in sample (Oñorbe et al. 2013) therefore we examine and make sure there is no contamination by low resolution particles inside the virial radius. Table 1 lists our sample halos, labeled m12b, m12f and m12i. Since there are two versions of each model, dark matter halo parameters are slightly different in each pair and all parameters are reported from the DMO model, unless otherwise mentioned (e.g., stellar mass, which is reported from SAGE). These numbers may also be different than what is reported in the original FIRE paper (Hopkins 2013) both because the simulation code is different and also different halo finders can characterize halos in different ways (Knebe et al. 2011) and cosmological simulations are chaotic in nature (Kandrup & Smith 1991; Keller et al. 2018; Genel et al. 2019). We produce 264 snapshots between redshift 20 and 0. Time increments are not linearly spaced but there is in average ~50Myr between successive snapshots. This is an optimum time resolution so we can follow the changes in the baryonic galaxy properties (see Section 4.7) without an overwhelming data overhead. We run the same initial condition and same configuration with both Gadget and CoSANG.

After we finished the simulations we perform the same analysis with the same configuration on the corresponding pairs.

5.2 Halo and galaxy growth

In section 4.4 we explained how SAGE calculates galactic parameters as halos grow and evolve in the simulation, using both the halo properties and merger history. In this section we discuss how m12b, m12f, m12i, and their associated galaxies grow their mass over time. Figure 3 shows the mass growth versus time. Colors represent different halos and circles and the star represent the major merger events. Merger events are defined as mass ratio of $M_{\text{ratio}} > 0.3$. The star is the merging of three massive halos with total mass of 1.5 times the mass of the main halo. The merger events are inferred by SAGE (Croton et al. 2016) and confirmed by tracking the merger tree.

m12b and m12f follow similar growth in their virial mass for most of the simulation and they both grow their mass earlier than m12i. m12b and m12f are early forming halos in our model while m12i is a late forming halo gaining much of its mass via a more violent merger that continues for a longer time therefore m12i catches up in mass to be a MW mass halo at $z \approx 0$. m12b has the last major merger, at $z \approx 2$, but it is a smaller event than m12i’s and does not dominate its mass growth in the same way. Note that halos do not truly decrease significantly in virial mass and the oscillation seen is due to the numerical problems with the halo finder (e.g., Knebe et al. 2011; More et al. 2011; Ludlow et al. 2019; Leroy et al. 2020).

The growth in stellar mass mostly mirrors the virial mass but in this case the similarity between m12b and m12f is not as significant. Stellar mass is less susceptible to fluctuations in the virial mass due to the semi-analytic prescription. Bulge mass to total mass ratio is shown in the bottom panel. In average all three galaxies reach $B/T \sim 0.6$.

The dynamical impact of the analytic baryonic potential versus dark matter potential is affected by the time and the scale of merger events. As we will see in section 5.3 a violent relaxation that occurs over a longer time scale on a late forming halo can wash out the importance of a baryonic potential on the final structure. Hydrodynamic models demonstrate similar diversity (Engler et al. 2021) and correlations between formation history and substructures (Rey & Starkenburg 2022).

5.3 Subhalo mass function and radial distribution

As we mentioned in section 3 the underlying dark matter kinematics is the key indicator to differentiate between different models. We study whether the baryonic galaxy affects the distribution of subhaloes (position, mass, number) in the model.

Every halo inside the virial radius is treated as a subhalo independent of its binding energy. We exclude all halos with fewer than 20 particles, corresponding to a subhalo mass limit of $M_{\text{vir}} > 7 \times 10^9 M_\odot$. Halos with less than 100 particles should be treated carefully.

We show Local Group observations in comparison with both models. The MW and M31 data are taken from McConnachie (2012); Simon (2019); Drlica-Wagner et al. (2020); Putman et al. (2021).

5.3.1 Mass function

Figure 4 shows the cumulative mass function for m12b (left), m12f (middle) and m12i (right). We examine the subhaloes mass abundance in both the DMO (solid blue) and CoSANG (solid red) simulations.

In m12b and m12f, the CoSANG simulation shows a reduced number of subhaloes for $M_{\text{subhalo}} < 10^9 M_\odot$. In other words, the baryonic potential of the galaxy can reduce the number of satellite galaxies and this effect is more significant on smaller satellites. At higher mass there is less significant differences and the most massive halos are slightly more massive in CoSANG compared to DMO.

We also plot the observed data for the MW (dashed black) and M31 (dashed gray). We used Putman et al. (2021) for most of the data. For galaxies without a reported mass, we estimate the dynamical mass using its effective radius and velocity dispersion. For a given velocity dispersion ($\sigma$), the mass interior to radius $r$ in an isothermal sphere is given by (Binney & Tremaine 2008)

$$M(r) = \frac{2\sigma^2 r}{G}.$$  \hspace{1cm} (17)

It is well known that DMO models produce a higher number of satellites compared to that around the MW (e.g. Klypin et al. 1999; Moore et al. 1999). Although observations can have an incomplete sample and the real number of satellites is higher than what is observed (Tollerud et al. 2008; Walsh et al. 2009; Hargis et al. 2014; Kim et al. 2018), baryonic processes can reduce the number of satellites and hydrodynamic models do not produce as many.
CoSANG simulations produce fewer subhaloes than DMO models, but still significantly more than the number of observed satellites around Local Group galaxies. This confirms the reduction of the number of subhaloes due to baryonic potentials.

Jahn et al. 2019 show number counts of subhaloes in simulations with baryons, within $0.2/\text{200}_{\text{vir}}$ ~ 70% of substructure depleted by the MW mass hosts, compared to simulations without baryons, and a factor of ~ 3.5 reduction (from DMO to baryonic runs) in the number of subhaloes at $v_{\text{max}} = 10$ km/s in MW mass hosts. Dynamical interactions with the disk can increase the likelihood of the presence of the subhaloes at closer distance however this can strip the subhaloes and decrease the number of subhaloes at this range.

The baryonic potential dominates at small radius and so is expected to have the largest effect there. Figure 5 shows that the depletion of subhaloes is indeed more significant within $\frac{1}{4} R_{\nu}$. CoSANG produces one more massive satellite than DMO and we see a few subhaloes are closer to the host halo in CoSANG compared to DMO.

m12i does not show as significant difference as two other models. The difference in the merger history and the environment can explain this difference between m12i and two other models. We will discuss these differences in more detail in section 5.6.

Observed satellites are selected by their stellar mass. Therefore, a better comparison between observed and simulated subhaloes should also incorporate stellar mass limits for the simulated sample. CoSANG $10^{4}$ (green) and CoSANG $10^{5}$ (brown) represent CoSANG with stellar mass cut $> 10^4 M_\odot$ and $> 10^5 M_\odot$ respectively. This cut-off reduces the lower end of the mass distribution significantly, in all three cases. More massive subhaloes are “too big to fail” (Boylan-Kolchin et al. 2011) and smaller subhaloes are not forming stars efficiently. This implies the stellar mass threshold has a significant impact. This reduced number is still much higher than hydrodynamic models.

We investigate the effect of a growing disk in time by following the evolution of the mass function at different redshifts. Figure 6 shows the distribution at two different redshifts ($z = 1.5$ top, $z = 0.3$ bottom). Although the simulation starts at redshift 20 and the galactic potential becomes part of the simulation as soon as galaxies form (which is $z \sim 13$, even though the stellar mass is not significant at high redshifts), still the three models are statistically similar up to redshift $z \approx 2$ when the stellar mass is large enough to produce a significant statistical difference. Since the main increase in the stellar mass is around this redshift, we see the difference becomes more pronounced after this at around $z \approx 1.5$ and the deviation between DMO and CoSANG increases as we progress in time. This indicates a correlation between the growth of a galactic potential and an increasing difference between DMO and CoSANG and not a random difference at different time-steps.

Observational comparisons are more robust when using velocities, which can be directly observed, rather than unobservable virial masses (McConnachie 2012; Fritz et al. 2020). Figure 7 shows the distribution of $v_{\text{max}}$ from DMO (blue) and CoSANG (red). Observed satellites of the MW (black dashed) and M31 (gray dashed) are shown for comparison. The observations are likely incomplete at lower $v_{\text{max}}$ (Tollerud et al. 2008; Hargis et al. 2014; Kim et al. 2018). The left panel shows the differential distribution and the right panel shows the cumulative distribution for m12b (top), m12f (middle) and m12i (bottom). $v_{\text{max}}$ peaks at smaller velocity in CoSANG compared to DMO and in general lies below the DMO distribution. CoSANG produces more lower $v_{\text{max}}$ satellites. Kelley et al. (2019) show that adding a disk potential into a DMO model results in a $v_{\text{max}}$ distribution below the DMO model, consistent with our results, but with a steeper slope (Kelley+19 shown as dotted black line). They sample 12 halos. Bullock+2000 (one halo) in dotted gray (Bullock et al. 2000) and W&S11 (six halos) in dotted light gray (Wadepuhl & Springel 2011) have distributions with shallower slopes. Different processes can affect the slope and hydrodynamic models produce shallower profiles (Wadepuhl & Springel 2011), therefore this difference in the slope could be due to underlying physical processes in the model.

While m12b and m12f demonstrate a similar distribution, nevertheless this difference is not very significant in m12i. We will discuss this difference in more detail in Section 5.6.

Observations are certainly incomplete at large radius, so the comparison at smaller radius is likely more useful. Figure 8 shows $v_{\text{max}}$ distribution at closer distance ($\frac{1}{4} R_{\nu}$). In this regime the CoSANG predictions bracket the observations well while the DMO predictions are higher.

Figure B1 shows the redshift evolution of $v_{\text{max}}$. As with the mass function, the difference between the $v_{\text{max}}$ distribution in CoSANG versus DMO becomes significant at $z<1.5$.

### 5.3.2 Radial profile

An analytic potential not only can change the distributions of subhalo masses but also their positions. In this section we study the radial distribution of subhaloes. Fig 9 shows the number of subhaloes enclosed in a given distance as a function of distance from the center, DMO (blue) and CoSANG (red) are the substructure distribution within the virial radius of m12b (left), m12f (middle) and m12i (right). Observed satellite galaxies of the MW (black dashed) and M31 (gray dashed) are shown for comparison. We see a uniform decrease in CoSANG relative to DMO at all radii, plus an increase at the closest distances.

The observed satellites of the Milky Way are quite consistent with CoSANG and lower than the DMO predictions at radii <50 kpc, beyond which incompleteness in the observations becomes a major issue (Tollerud et al. 2008; Hargis et al. 2014; Kim et al. 2018; Samuel et al. 2020).

Like the mass distribution, the redshift evolution of the radial distribution starts at $z=1.5$ and grows with time (Figure B2).

A comparison between a DMO and DMO+disk; in PhatELVIS, shows a similar trend (Kelley et al. 2019). Carlsten et al. (2020) compare the radial distribution of subhaloes in variety of models (DMO, DMO+disk, hydro, box and zoom-in) with observations and show subhaloes of the MW mass galaxies are more centrally

| ID | Mass resolution [$M_\odot h^{-1}$] | $M_{\nu} [M_\odot h^{-1}]$ | $R_{\nu} [h^{-1} \text{kpc}]$ | DMO subhaloes | CoSANG subhaloes | $M_{\nu} [M_\odot]$ | $R_{\nu} [\text{kpc}]$
|----|---------------------------------|--------------------------|--------------------------|---------------|------------------|-----------------|---------------|
| m12b | $2.4 \times 10^9$ | $6.6 \times 10^{11}$ | 142 | 1635 | 1106 | 6.1 $\times 10^{10}$ | 3.2 |
| m12f | $2.4 \times 10^9$ | $7.8 \times 10^{11}$ | 150 | 2834 | 2052 | 7.8 $\times 10^{10}$ | 3.8 |
| m12i | $2.4 \times 10^9$ | $5.5 \times 10^{11}$ | 134 | 1584 | 1733 | 5.0 $\times 10^{10}$ | 2.8 |
Figure 3. Growth of the virial mass (top) and stellar mass (middle) and bulge to total stellar mass ratio (bottom) for individual halos with CoSANG. The circles and star represent major merger events. m12b and m12f have similar halo growth, while the growth and major mergers in m12i occur slightly later. The star is the most violent event of all major merger events. Stellar mass follows a similar pattern but the similarity between m12b and m12f is not as significant as for the virial mass.

5.4 Halo shape

We adapt the same approach as Zemp et al. (2011) to measure the triaxial dark matter halo shape. We use the second moment of the mass distribution as shape:

\[
S_{ij} = \frac{\int k m_k (\mathbf{r}_k)_i (\mathbf{r}_k)_j \, dk}{\int k m_k \, dk}
\]

(18)

Where \((\mathbf{r}_k)_j\) denotes the j-component of the position vector of the kth particle.

For each radial bin we calculate the second moment of the mass distribution and we find its eigenvalues and eigenvectors. Eigenvalues are proportional to the square of axis lengths; we denote them from the longest to the shortest as \(a–b–c\) with associated eigenvectors \(\mathbf{A}, \mathbf{B}, \mathbf{C}\) representing the direction of each axis. In the next iteration for this bin we change the volume from which particles are selected to be an ellipsoidal shell whose shape and direction are consistent with the previous iteration, and iterate until we converge or until we reach the iteration limit. One important problem with determining the shape is the treatment of the subhaloes, which we choose to exclude.

We use the average radius of the bins as corresponding radii and we study the evolution of the shape as a function of the distance from the host.

Figure 10 shows the axis ratios \(b/a\) (red) and \(c/a\) (blue) in both DMO (dashed) and CoSANG (solid) for m12b (left), m12f (middle) and m12i (right). In m12b the difference between two models are more significant at distances closer than 100kpc and beyond this distance the axes are almost the same. In m12f we see axis ratios are lower in CoSANG. The difference in axis ratios in mid-radii is not as significant as at small and large radii. In m12i the difference is significant everywhere within the virial radius. In all three halos, CoSANG produces smaller axis ratios so it is less spherical than DMO. This could be mainly due to the existence of a disk potential at the center. This is not consistent with similar studies with hydrodynamic model that produce rounder halos compared to DMO models. One possible explanation could be the dynamical effect of the feedback that is stronger than the dynamical effect of the disk. Feedback is accounted for indirectly in CoSANG, via SAGE, but not directly. This could be an interesting update to study the dynamic effect of the feedback in future work.

One interesting result is the significant difference between DMO and CoSANG in m12i. Although this late forming halo with violent relaxation does not show a significant difference in its sub-
Figure 4. Cumulative mass function of subhaloes in halo m12b (left), m12f (middle) and m12i (right). The blue line represents DMO, the red line represents CoSANG, the black dashed line is the MW, and the gray dashed line is M31. CoSANG $10^4$ (green) and CoSANG $10^6$ (brown) represent CoSANG with stellar mass cut $\leq 10^4 M_\odot$ and $\geq 10^6 M_\odot$ respectively. This figure contains all subhaloes within the virial radius. In general the CoSANG simulations contain fewer subhaloes than the DMO simulations. m12b and m12f display a more significant difference; m12i, which has a different merger history, shows no difference between CoSANG and DMO simulations. CoSANG shows reduced substructure but not enough on its own to explain the missing satellite problem.

Figure 5. Cumulative mass function of subhaloes at $r < \frac{1}{2} R_\nu$ in halo m12b (left), m12f (middle) and m12i (right). The blue line represents DMO, the red line represents CoSANG, the black dashed line is the MW and the gray dashed line is M31. The red shaded region is the mass resolution limit. The difference between DMO and CoSANG is more significant at smaller radii, especially for m12b. m12i does not show any difference.

halo distribution, its shape shows a significant difference between DMO and CoSANG.

The shortest axis being reduced in CoSANG may reflect the orientation of the angular momentum of the halo, which sets the direction of the disk potential (see Section 4.5). We can examine this in Figure 11 where we compare the direction of each axis to that of the angular momentum vector. We use the total angular momentum of the halo to compare the orientation of the shape with the angular momentum and consequently the galactic disk. We plot each component separately ($\hat{A} \cdot \hat{L}$ in blue, $\hat{B} \cdot \hat{L}$ in red, $\hat{C} \cdot \hat{L}$ in green) in DMO (dashed) and in CoSANG (solid) for both m12b (left), m12f (middle) and m12i (right).

For m12b and m12f in CoSANG, the minor axis is aligned with the angular momentum in DMO, as seen in other studies (e.g. Bett et al. 2010; Teklu et al. 2015; Obreja et al. 2022) CoSANG modifies the inner regions and makes them align more with the intermediate axis. Similar weak alignments and a large scatter in the distribution of these alignments have been studied in different models (Bett et al. 2010; Vera-Ciro et al. 2011; Schneider et al. 2012; Bryan et al. 2013) which suggest the angular momentum...
Redshift evolution of cumulative mass function of subhaloes in m12b (left), m12f (middle) and m12i (right). Blue lines represent DMO and red lines represent CoSANG. This figure includes all subhaloes within the virial radius. The shaded red region is our mass limit. The difference is not significant at z>1.5 and the two models are statistically similar at higher redshifts. However, the differences increase with time as the galaxy grows and the dynamical effects have time to act on substructures.

and subsequently the disk orientation can play an important role in determining the shape. On the other hand, in m12i the alignment switches in CoSANG from being predominantly along the minor axis to being predominantly along the major axis.

CoSANG produces more prolate and less spherical shapes than DMO. This demonstrates that the non-linear effects of baryonic physics can result in more complicated differences, and many earlier studies have found correlations between the angular momentum and environment, assembly history and characteristics of the halo (e.g. Gao et al. 2005; Faltenbacher & White 2010; Libeskind et al. 2012; Schneider et al. 2012; Sharma et al. 2012; Trowland et al. 2013; Genel et al. 2015; Zjupa & Springel 2017).

5.5 Halo twists
We can study internal twists in the halo shape to understand if different layers respond to the central potential or if the inter-layer interaction can dominate the kinematics. We compare the orientation in each bin with the first bin (residual orientation). Figure 12 shows the results in DMO (dashed) and CoSANG (solid) for m12b (left), m12f (middle) and m12i (right) for the three axis vectors (in colors) considered previously. In both m12b and m12i, we do not see a significant difference between DMO and CoSANG and in both halos all layers have relatively well-aligned shape orientations with no sudden changes. The major axis A has the highest self-alignment but B and C experience small changes. m12f, on the other hand, demonstrates a significant twist in the shape in CoSANG.

5.6 Diversity of substructures
CoSANG allows us to understand the impact of the baryonic potential on substructures. m12b and m12f demonstrate differences between DMO and CoSANG as one might expect; however m12i raises the question why it shows a different behavior to the two
Figure 7. $V_{\text{max}}$ distribution (left) and cumulative $V_{\text{max}}$ distribution (right) for subhaloes within the virial radius of m12b (top), m12f (middle) and m12i (bottom). DMO is in blue and CoSANG is in red. CoSANG $10^4$ (green) and CoSANG $10^6$ (brown) represent CoSANG with stellar mass cut $> 10^4 M_\odot$ and $> 10^6 M_\odot$ respectively. Observed satellites of the MW (black dashed) and M31 (gray dashed) are shown for comparison. For a given $V_{\text{max}}$, the CoSANG simulations contain fewer subhaloes than DMO and $V_{\text{max}}$ peaks at a smaller value. Kelley+19 (dotted black line) (Kelley et al. 2019), Bullock+2000 (dotted gray (Bullock et al. 2000) and W&S11 (dotted light gray) (Wadepuhl & Springel 2011) show the slope from other studies with 12,1,6 halos in their sample respectively.
Figure 8. $V_{\text{max}}$ distribution (left) and cumulative $V_{\text{max}}$ distribution (right) for subhaloes at $r < \frac{1}{4} R_v$ of m12b (top), m12f (middle) and m12i (bottom). DMO is in blue and CoSANG is in red. Observed satellites of the MW (black dashed) and M31 (gray dashed) are shown for comparison. For a given $V_{\text{max}}$ CoSANG simulations contain fewer subhaloes compared to DMO.
Figure 9. Cumulative radial distribution of subhaloes of m12b (left), m12f (middle) and m12i (right). This shows all subhaloes within the virial radius and DMO is in blue and CoSANG is in red. Observed satellites of the MW (black dashed) and M31 (gray dashed) are shown for comparison. CoSANG 10^4 (green) and CoSANG 10^6 (brown) represent CoSANG with stellar mass cut > 10^3 M_☉ and > 10^6 M_☉ respectively. The number of subhaloes in CoSANG is reduced compared to DMO at most distances, though it shows an increase in the very innermost region. CoSANG appears quite consistent with the MW data within 50 kpc, beyond which the observational data is likely quite incomplete.

Figure 10. Shapes as function of radial distance from the center of mass of the halo in m12b (left), m12f (middle) and m12i (right). b/a is in red, c/a is in blue, and DMO and CoSANG are shown as the dashed and solid lines respectively. There is a difference between DMO and CoSANG for all halos, always in the sense that the axis ratios are smaller in CoSANG, but the radial dependence varies between halos.

other halos? We believe this is an interesting outcome supporting the importance of a self-consistent interactive hybrid models like CoSANG. Although all halos are in the same mass range, they experience different formation histories. Merger history and environment can change the substructures and violent relaxation can dominate other forces. A comparison between different models including full physics hydrodynamic simulations shows a wide range of substructures even for the same mass host (Engler et al. 2021). Rey & Starkenburg (2022) demonstrated using genetically modified DMO models that early forming and late forming halos can exhibit different structure in the dark matter particles that they tagged with stellar populations.

As discussed in section 5.2, m12i is a late forming halo and has a different merger history compared to m12b and m12f. This difference is reported in other similar studies of simulations that use the same initial conditions, where capturing the unusual accretion history of m12i becomes a challenge (Cunningham et al. 2021). SAMs are sensitive to the merger history of the halo and this difference can...
Figure 11. Shape axis vs angular momentum orientation in m12b (left), m12f (middle) and m12i (right). Individual shape axis (A in blue, B in red, C in green) alignment with the angular momentum both in DMO (dashed) and CoSANG (solid). In m12b and m12f CoSANG and DMO halos have similar alignment at distances larger than 40 kpc, while within inner 40 kpc, we see more alignment between the disk and B and slightly less alignment between the disk and C. Interestingly m12i displays a very different behavior: its disk is more aligned with A in CoSANG while it is aligned with C in DMO. The B and C mixing at R>125 kpc is due to degeneracy between the axes in this almost-prolate region.

Figure 12. Shape axis residual in m12b (left), m12f (middle) and m12i (right). The residual is the orientation of the axis at each radial bin (A in blue, B in red, and C in green) relative to the orientation of that axis in the innermost bin in both DMO (dashed) and CoSANG (solid). The haloes have a relatively consistent orientation at all distances. m12b and m12i have insignificantly more twists in CoSANG. m12f demonstrates a significant twist.

leave their fingerprint on the final substructure. Therefore we can not ignore this assembly history, and hence any hybrid model should include baryonic potentials that reflect not only the halo abundance but also the associated environment and history. CoSANG has the advantage of including all of these considerations.

Such a diversity is reported in hydrodynamic models (e.g., Samuel et al. 2020, Engler et al. 2021) but DMO models do not show as significant a diversity (Kelley et al. 2019). Although there are differences in the sample size in different studies and different methods we may use to identify distributions, we can safely hypothesize the environment and formation history may explain most of the differences between different models and similar studies. We choose these galaxies based on their mass at $z=0$ but they don’t necessarily follow the same evolution as the MW and their ultimate properties may fail to capture these differences. This is one of the strongest features of CoSANG, that we implement a full SAM. This is a significant difference compared to similar studies. CoSANG includes different components of the galaxy and the SAM is very sensitive not only to the halo abundance but also the merger tree, so even similar halo abundances can produce different galaxy pa-
rameters. This makes CoSANG a very powerful toolkit to create large samples of galaxies having a diversity in properties capturing individual galaxy formation histories.

6 CONCLUSIONS

We present CoSANG (Coupled Semi-Analytic N-body Galaxies), a new cosmological model developed for stellar halo simulations, but which has many additional applications. In this paper, we use it to study the structure of dark matter halos of Milky Way mass galaxies. CoSANG is a hybrid model with an N-body core and a semi-analytic model (SAM) coupled at every time-step. CoSANG allows for more accurate dynamics than a pure dark matter simulation due to the inclusion of a potential corresponding to the baryonic component of the galaxy and these dynamics have significant effect on substructures. CoSANG has been tested using a variety of scenarios from isolated orbit models to large scale cosmological simulations.

We performed zoom-in simulations of three Milky Way (MW) mass halos and analyzed their substructures in both dark matter only (DMO) and CoSANG cases. We find the following:

- The dynamical effect of a baryonic potential is significant on the inner structure of the halo.
- CoSANG produces a smaller number of subhaloes compared to the DMO model. This is consistent with similar studies indicating the galactic disk can reduce the number of subhaloes. The baryonic potential can both enhance the tidal field to disrupt substructure, and also bring them in to smaller radius. Thus we see the number of subhaloes has decreased in CoSANG and also slightly enhanced at small radii. We still see a higher number of subhaloes compared to hydrodynamic models. This suggest we will need more than baryonic potentials (such as feedback force) to reproduce similar results as hydrodynamic models. Adding a stellar mass threshold reduces the number of satellites significantly, especially at lower mass end.
- N-body models produce many subhaloes. At radii where observational samples of Milky Way satellites are expected to be mostly complete, the reduced number of CoSANG subhaloes provides a reasonably good match, but that it still predicts more than the number of M31 satellites.
- The SAM SAGE predicts galaxies start forming at very high redshift (z ∼ 13), and there is no significant difference between CoSANG and DMO until z ∼ 1.5 but this difference then grows with time.
- Although adding a galactic potential can result in differences between DMO and CoSANG, our MW analogue m12i, that formed later and had a more significant major merger does not exhibit much difference between DMO and CoSANG, unlike the others (m12b, m12f). All three halos have similar masses but different merger histories and different environments can change the significance of this variation from one halo to another.
- CoSANG produces more stretched (more prolate) shapes compared to DMO. Hydrodynamic models produce more oblate shapes. This may suggest the galactic potential is not the main factor determining the final shape of a halo and feedback may play a more important role in shape of the halo. This could be investigated in future update of CoSANG to include the dynamical effect of the feedback. The minor axis is aligned with the angular momentum of the halo in both models but this alignment is slightly different at closer distances in CoSANG where we see more alignment between the intermediate axis and angular momentum.

- The total number of subhaloes in CoSANG is lower than DMO and their radial separation is more centrally concentrated.

A hybrid model like CoSANG with an empirical baryonic potential is computationally efficient. Although this model is not a replacement to a full physics hydrodynamic simulation, it is a complementary model and a powerful tool to study galactic dynamics and the stellar halo.

CoSANG has been used here to study dark matter kinematics. But CoSANG is also capable of producing stellar halo populations through particle tagging. Its use for stellar halo kinematics and stellar halo populations will be presented in an upcoming paper (Talei et al., in preparation).

ACKNOWLEDGEMENTS

The authors are sincerely grateful for every help we received throughout the development of CoSANG. Astronomy faculty members at the University of Alabama were supportive during entire development process. We are also immensely grateful to Monica Valluri (University of Michigan) for her contribution to CoSANG and comments on an earlier version of the manuscript. We also acknowledge useful conversations with galaxy theory group at Steward observatory. Especially Gurtina Besla, Peter Behroozi and Nicolas Garavito Camargo have offered many insightful comments. We gratefully acknowledge the support of XSEDE Campus Champions program.

We ran all cosmological simulations using UAHPC (University of Alabama High Performance Computing) and TACC (Texas Advanced Computing Center) through XSEDE allocations (TG-AST200026 and TG-PHY200076) and OZ-STAR (Swinburne supercomputing center). Most post-processing and analysis was done with UAHPC.

SOFTWARE ACKNOWLEDGEMENT

All codes we use are publicly available at this URL: http://www.github.com/stalei

Many of these codes are developed and used for the first time. The authors respect and encourage reproducible results as an important tool for improvement. All first time codes are under public domain licence and using and improving them with reference is encouraged. We also acknowledge the use of the following packages: Galstep: https://github.com/ruggiero/galstep, numpy: Harris et al. 2020, Rockstar: Behroozi et al. 2012, Scipy: Jones et al. 2001, Splash: Price 2007, yt: Turk et al. 2011.

DATA AVAILABILITY

All data used in this work is produced with CoSANG. The simulation output will be public as soon as the full sample is complete. Data is available upon request.
APPENDIX A: ORBIT TEST

One of the most important tests of the galactic potential embedded in the code is the orbit test. We test orbits both with CoSANG and with an independent orbit integrator. The initial condition is produced manually and consists of a handful of particles with precise velocities and positions. We configured CoSANG for non-cosmological simulation. We put test particles with different initial velocities and positions and we let them evolve under this potential. The orbit of these test particles are well-understood (Binney & Tremaine 2008). These tests confirm that the new analytic potential is consistent with known solutions. We performed multiple orbit tests; Figure A1 shows two orbits for a particle orbiting around the center at the expected circular velocity without any perpendicular component (blue) and with a velocity component perpendicular to the disk (red). We can see the flat rotation and a stable oscillating orbits. We recover the same results with the orbit integrator and with full CoSANG, both consistent with analytic orbits.

APPENDIX B: REDSHIFT EVOLUTION

The cumulative $V_{\text{max}}$ (B1) and the radial abundance of subhaloes (B2) at two different redshifts for both DMO (blue) and CoSANG (red) and all three models m12b (right), m12f (middle) and m12i (left). There is no significant difference between DMO and CoSANG at redshift 1.5 and higher. The difference increases after redshift 1.5.

This paper has been typeset from a \TeX/\LaTeX\ file prepared by the author.
Figure B1. The redshift evolution of the cumulative $V_{\text{max}}$ distribution. Redshift 1.5 (top) and redshift 0.3 (bottom) for m12b (left), m12f (middle) and m12i (right). The difference in the distribution is negligible at redshift 1.5 and becomes more significant by redshift 0.3.
Figure B2. Redshift evolution of the radial distribution of subhaloes within the virial radius of m12b (left), m12f (middle) and m12i (right) in DMO (blue) and CoSANG (red) at redshift 1.5 (top) and redshift 0.3 (bottom). CoSANG and DMO are almost identical at redshift 1.5. In m12b the deviation between the models increases as the galaxy grows. In m12f and m12i this deviation is less significant.