

## Introducing the DREAMS Project: DaRk mattEr and Astrophysics with Machine learning and Simulations

JONAH C. ROSE,<sup>1,2</sup> PAUL TORREY,<sup>3</sup> FRANCISCO VILLAESCUSA-NAVARRO,<sup>2,4</sup> MARIANGELA LISANTI,<sup>2,5</sup> TRI NGUYEN,<sup>6,7</sup> SANDIP ROY,<sup>5</sup>  
KASSIDY E. KOLLMANN,<sup>5</sup> MARK VOGELSBERGER,<sup>6,7</sup> FRANCIS-YAN CYR-RACINE,<sup>8</sup> MIKHAIL V. MEDVEDEV,<sup>4,9,10,11,12</sup> SHY GENEL,<sup>2,13</sup>  
DANIEL ANGLÉS-ALCÁZAR,<sup>14,2</sup> NITYA KALLIVAYALIL,<sup>3</sup> BONNY Y. WANG,<sup>15</sup> BELÉN COSTANZA,<sup>16</sup> STEPHANIE O'NEIL,<sup>6</sup> CIAN ROCHE,<sup>6</sup>  
SOUMYODIPTA KARMAKAR,<sup>8</sup> ALEX M. GARCIA,<sup>3</sup> RYAN LOW,<sup>11</sup> SHURUI LIN,<sup>17,18</sup> OLIVIA MOSTOW,<sup>3</sup> AKAXIA CRUZ,<sup>2,5</sup> ANDREA CAPUTO,<sup>19</sup>  
ARYA FARAH,<sup>20</sup> JULIAN B. MUÑOZ,<sup>21</sup> LINA NECIB,<sup>6,7</sup> ROMAIN TEYSSIER,<sup>4</sup> JULIANNE J. DALCANTON,<sup>2,22</sup> AND DAVID SPERGEL<sup>2</sup>

<sup>1</sup>*Department of Astronomy, University of Florida, Gainesville, FL 32611, USA*

<sup>2</sup>*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*

<sup>3</sup>*Department of Astronomy, University of Virginia, Charlottesville, VA 22904, USA*

<sup>4</sup>*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*

<sup>5</sup>*Department of Physics, Princeton University, Princeton, NJ 08544, USA*

<sup>6</sup>*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

<sup>7</sup>*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, MA 02139, USA*

<sup>8</sup>*Department of Physics and Astronomy, University of New Mexico, 210 Yale Blvd NE, Albuquerque, NM 87106, USA*

<sup>9</sup>*Institute for Advanced Study, School for Natural Sciences, Princeton, NJ 08540*

<sup>10</sup>*Institute for Theory and Computation, Harvard University, Cambridge, MA 02138*

<sup>11</sup>*Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045*

<sup>12</sup>*Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA 02139*

<sup>13</sup>*Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA*

<sup>14</sup>*Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA*

<sup>15</sup>*McWilliams Center for Cosmology and Astrophysics, Carnegie Mellon University, Pittsburgh, PA 15213*

<sup>16</sup>*Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Observatorio Astronómico, Paseo del Bosque, B1900FWA La Plata, Argentina*

<sup>17</sup>*Department of Astronomy, School of Physical Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China*

<sup>18</sup>*CAS Key Laboratory for Researches in Galaxies and Cosmology, School of Astronomy and Space Science, University of Science and Technology of China, Hefei, Anhui 230026, China*

<sup>19</sup>*Theoretical Physics Department, CERN, 1211 Geneva 23, Switzerland*

<sup>20</sup>*Departments of Statistics and Data Science, University of Texas at Austin, Austin, TX 78757, USA*

<sup>21</sup>*Department of Astronomy, University of Texas at Austin, Austin, TX 78757, USA*

<sup>22</sup>*Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA*

### ABSTRACT

We introduce the DREAMS project, an innovative approach to understanding the astrophysical implications of alternative dark matter models and their effects on galaxy formation and evolution. The DREAMS project will ultimately comprise thousands of cosmological hydrodynamic simulations that simultaneously vary over dark matter physics, astrophysics, and cosmology in modeling a range of systems—from galaxy clusters to ultra-faint satellites. Such extensive simulation suites can provide adequate training sets for machine-learning-based analyses. This paper introduces two new cosmological hydrodynamical suites of Warm Dark Matter, each comprised of 1024 simulations generated using the AREPO code. One suite consists of uniform-box simulations covering a  $(25 h^{-1} \text{ M}_\odot)^3$  volume, while the other consists of Milky Way zoom-ins with sufficient resolution to capture the properties of classical satellites. For each simulation, the Warm Dark Matter particle mass is varied along with the initial density field and several parameters controlling the strength of baryonic feedback within the IllustrisTNG model. We provide two examples, separately utilizing emulators and Convolutional Neural Networks, to demonstrate how such simulation suites can be used to disentangle the effects of dark matter and baryonic physics on galactic properties. The DREAMS project can be extended further to include different dark matter models, galaxy formation physics, and astrophysical targets. In this way, it will provide an unparalleled opportunity to characterize uncertainties on predictions for small-scale observables, leading to robust predictions for testing the particle physics nature of dark matter on these scales.

## 1. INTRODUCTION

The DaRk mattEr and Astrophysics with Machine learning and Simulations (DREAMS<sup>1</sup>) project aims to disentangle how variations in both dark matter (DM) and baryonic physics impact galaxy formation. Integral to this effort is the development and training of Machine Learning models that have the capacity to identify subtle impacts from changing DM model assumptions in the face of uncertain baryonic physics. The project will ultimately comprise thousands of simulations, spanning Cold Dark Matter (CDM) and more complicated dark-sector models (for a review, see Bertone & Hooper 2018) while simultaneously incorporating uncertainties in model predictions by varying baryonic feedback across the simulations. Different DM scenarios will be simulated for a range of targets—from clusters to ultra-faint satellites—providing an opportunity to study the impact of fundamental particle physics interactions on galactic and sub-galactic scales. This paper introduces the DREAMS project and provides some case studies to demonstrate its use in machine learning applications.

The CDM paradigm has proven to be remarkably successful on large cosmological scales (Springel et al. 2005; Reid et al. 2010; Alam et al. 2017; Springel et al. 2018), and the current frontier is to test its validity on galactic and sub-galactic scales. In this paradigm, the DM is typically collisionless and interacts primarily through gravity. However, recent theoretical work highlights that DM can live in a “dark sector,” with multiple particles and/or new forces that mediate non-gravitational interactions (Bertone & Tait 2018). Such dark-sector models can reproduce cosmological observables and yield the correct abundance of DM today, while simultaneously affecting the internal structure of galaxies as they evolve (for a review, see Arun et al. 2017).

The potential degeneracy that arises from modifications to the internal structure of galaxies has led to an ongoing debate on several small-scale “tensions” where observations appear to conflict with CDM expectations (for a review, see Sales et al. 2022). Classic examples include the diversity (Kuzio de Naray et al. 2010; Oman et al. 2015), missing-satellites (Moore et al. 1999; Klypin et al. 1999), too-big-to-fail (Boylan-Kolchin et al. 2011, 2012; Garrison-Kimmel et al. 2014; Tollerud et al. 2014; Papastergis et al. 2015), satellite planes (Pawlowski et al. 2014; Pawlowski 2018), and quiescent fractions (Hausammann et al. 2019; Samuel et al. 2022) problems. In some of these cases, the apparent tensions can be explained by an incomplete knowledge of baryonic feedback (Governato et al. 2010; Brooks et al. 2013) and/or observational biases (Font et al. 2022; Sawala et al.

2023). For example, the missing-satellites problem, where many more satellites were found in cosmological simulations than observed around the Milky Way (MW), is now considered solved after self-consistent stellar feedback was implemented in cosmological simulations (Garrison-Kimmel et al. 2019; Engler et al. 2021) and observational completeness was accounted for (Nadler et al. 2020).

One of the key ingredients of the DREAMS project is the treatment of astrophysical uncertainties that arise from baryonic feedback processes. Due to the limiting resolution of current hydrodynamic cosmological simulations, the astrophysics of stellar and black hole evolution is not emergent and must be implemented through phenomenological models. While some models (e.g. Hopkins et al. 2018; Marinacci et al. 2019) are progressing toward more explicit implementations of the physical processes that drive galaxy formation, any given “sub-grid” prescription depends on a set of parameters that remain uncertain, even if constrained to some extent by available observations. For a review of commonly used subgrid models and their components, see Vogelsberger et al. (2020). Because energy injection due to Supernova (SN) and Active Galactic Nuclei (AGN) feedback plays an important role in redistributing the matter at the center of a halo (Governato et al. 2010; Teyssier et al. 2011; Peirani et al. 2017, 2019; Gebhardt et al. 2023), large uncertainties in feedback modeling can translate into large uncertainties on certain DM properties. Accounting for these astrophysical systematics is therefore of critical importance to produce robust predictions for these properties. The DREAMS project will include thousands of simulations that span the uncertainty in alternative DM and galaxy formation models to catalyze the theory needed to utilize galaxies as particle physics laboratories. With this dataset, the DREAMS project will work toward the ultimate goal of testing CDM at the galaxy formation scale or discovering its alternate.

Dedicated cosmological hydrodynamic simulations exist for a small number of CDM alternatives (e.g. Vogelsberger et al. 2014b; Robles et al. 2017; Lovell et al. 2018; Robertson et al. 2019; Sameie et al. 2021; Shen et al. 2022a; Kulkarni et al. 2022; Rose et al. 2023; Brown et al. 2024; Correa et al. 2024), but—due to their limited number—do not adequately account for cosmic variance, uncertainties in baryonic modeling, or the extent of available DM parameter space. The DREAMS project aims to adequately span this large multidimensional parameter space. Given the breadth of simulation data that will be available, the non-linear nature of structure formation, and the apparent degeneracies between the dark sector and baryonic physics, the project lends itself to statistical applications, such as machine learning. The CAMELS project (Villaescusa-Navarro et al. 2021a), which contains thousands of CDM simulations where astrophysical and cosmological parameters have been varied, pro-

<sup>1</sup> <https://dreams-project.readthedocs.io>  
<https://www.dreams-project.org>

vides a successful demonstration of the analyses that can ultimately be performed with DREAMS. For example, [Ni et al. \(2023\)](#) showed that their neural networks can disentangle 28 simultaneously-varied parameters to make accurate constraints on cosmology and astrophysics. Additionally, other works within the CAMELS project have shown that a variety of machine learning models can be used to constrain cosmological parameters while marginalizing over uncertainties in baryonic physics ([Villaescusa-Navarro et al. 2021b](#); [Villanueva-Domingo & Villaescusa-Navarro 2022](#); [Villaescusa-Navarro et al. 2022a](#); [de Santi et al. 2023a](#)). By exploring alternative DM models and extending the simulations down to smaller scales, the DREAMS project will open new avenues of exploration for using machine learning methods to infer DM properties based on small-scale effects.

This paper presents the first two hydrodynamical simulation suites that have been completed for the DREAMS project. For simplicity, we choose to start with thermal-relic Warm Dark Matter (WDM; [Hogan & Dalcanton 2000](#); [Sommer-Larsen & Dolgov 2001](#)) where the initial matter power spectrum is suppressed below the free-streaming scale of the DM particle. This model introduces only one additional simulation parameter (i.e., the WDM particle mass) relative to CDM. The suppression in small-scale power makes WDM highly predictive. Current  $2\sigma$  constraints from a joint analysis of strong gravitational lensing and satellite properties rule out masses below 9.7 keV ([Nadler et al. 2021a](#)). Another joint analysis by [Enzi et al. \(2021\)](#), which incorporates  $\text{Ly}\alpha$ -forest measurements, constrains WDM masses to  $> 6.0$  keV. Other constraints have also been set at 6.5 keV through observations of satellite properties ([Nadler et al. 2021b](#)), 5.6 keV through strong gravitational lensing ([Hsueh et al. 2020](#)), and 5.3 keV through measurements of the  $\text{Ly}\alpha$ -forest ([Iršič et al. 2017](#)).

We present results for both a uniform-box and MW zoom-in WDM simulation suite. Each suite contains 1024 distinct simulations over which parameters associated with SN and AGN feedback have been varied within the IllustrisTNG (TNG; [Pillepich et al. 2018a](#); [Weinberger et al. 2018](#)) galaxy formation model. Notably, this is the largest zoom-in suite created to date for any DM model and provides a unique sample of MW-like classical satellites to study in a WDM cosmology.

We perform two separate case studies with these suites. The first utilizes emulation on the zoom-in suite to investigate how WDM and astrophysical parameters affect satellite counts around MW-mass hosts. We find that the machine-learning-based analysis can disentangle how DM and astrophysics parameters independently affect satellite counts. The second infers WDM particle properties from field-level data with CNNs using images of projected density fields from uniform-box simulations that vary DM, baryon, and cosmological properties. With this dataset, our analysis marginal-

izes over baryonic and cosmological uncertainties to infer the WDM particle mass from a range of astrophysical properties.

Moving forward, the DREAMS project will grow to include additional suites for other DM models, including examples with multiple subspecies of DM and/or novel DM interactions. Within this framework, we aim to:

- understand the dependence of galaxy formation and galaxy properties on variations in the dark sector;
- identify new observables that can constrain properties of DM while accounting for uncertainties in cosmology and astrophysics;
- efficiently explore large swaths of dark-sector parameter space to set constraints and identify promising regions for detectability;
- train machine-learning models to infer DM properties while marginalizing over cosmological and astrophysical uncertainties.

This paper is organized as follows. Section 2 outlines the simulation design and specifications within the DREAMS project. The next two sections, Sections 3 and 4, outline the two different applications for the WDM suites that have been completed. Section 5 outlines the possible extensions of this framework. We conclude in Section 6. The appendices provide additional information for the methods and models included in this paper. Appendix A outlines the procedure used to generate the zoom-in suite consisting of 1024 unique MW-mass galaxies. Appendix B details the different alternative DM models presented in this paper. Finally, Appendix C provides information on the machine learning architectures, training procedures, and data used in the analysis.

## 2. SIMULATION FRAMEWORK

The DREAMS project is comprised of separate suites of cosmological simulations— either uniform boxes or zoom-ins. Uniform boxes have the advantage of providing large statistics with millions of galaxies but at the cost of poor resolution. Zoom-in simulations focus on fewer galaxies than uniform boxes but at much higher resolution for a similar computational cost. Including multiple simulation types allows one to focus on different observational probes of galactic structure, while the large number of simulations in each suite helps to capture cosmic variance. Each hydrodynamical simulation in the suite also has a different set of baryonic feedback parameters. These variations allow one to incorporate uncertainties in the astrophysical prescription into the final results, similar to what is done in CAMELS ([Villaescusa-Navarro et al. 2021a](#)). Table 1 summarizes the WDM hydrodynamical suites used in this paper, as well as currently available or forthcoming suites that are also part of the DREAMS project.

Dark Matter Prescription	# of Sims	Simulation Type	Baryon Prescription	Cosmology	Mass Resolution [ $h^{-1}\Omega_m M_\odot$ ]	Spatial Resolution [ $h^{-1}$ pc]	Notes
Cold Dark Matter	1024	Milky Way	N-body	Varied	$4.0 \times 10^6$	305	Forthcoming
	1024	Milky Way	TNG	Varied	$4.0 \times 10^6$	305	Forthcoming
Warm Dark Matter	200	Uniform Box	N-body	Fixed	$3.3 \times 10^7$	1000	<a href="#">Rose et al. (2024)</a>
	1000	Uniform Box	N-body	Varied	$2.6 \times 10^8$	1000	<a href="#">Rose et al. (2024)</a>
	1024	Uniform Box	TNG	Varied	$2.6 \times 10^8$	1000	This paper
	1024	Milky Way	TNG	Fixed	$4.0 \times 10^6$	305	This paper

**Table 1.** List of simulation parameters for each suite in the DREAMS project. *Col 1:* The name of the dark matter model included in the suite. *Col 2:* The number of unique simulations in each suite, each with its own random initial density perturbations as well as dark matter and/or baryonic physics parameters (if relevant). *Col 3:* The type of simulation, either uniform box or Milky Way zoom-in. *Col 4:* The galaxy formation model that is included in the simulations, if any. *Col 5:* Whether the cosmological parameters  $\Omega_m$  and  $\sigma_8$  are varied within the suite. *Col 6:* The dark matter mass resolution for each of the simulation suites divided by  $\Omega_m$ . Note that dividing by  $\Omega_m$  makes the mass resolutions appear larger than those typically quoted in the literature. *Col 7:* The gravitational softening for each simulation suite at  $z = 0$ . The N-body WDM suites were studied in [Rose et al. \(2024\)](#), while the TNG WDM suites are the focus of this paper. The next suites planned as part of the DREAMS effort will focus on CDM, denoted above as “forthcoming.” Section 5 discusses plans for other extensions to the DREAMS project.

For the suites considered in this work, the simulations are advanced from  $z = 127$  to  $z = 0$  by the moving-mesh code AREPO ([Springel 2010](#); [Springel et al. 2019](#); [Weinberger et al. 2020](#)). AREPO uses a TreePM grid ([Bagla 2002](#)) to solve gravity with a comoving spatial softening for all particles to reduce the numerical noise of close-body interactions. Hydrodynamics are implemented using a pseudo-Lagrangian moving-mesh approach, which utilizes Voronoi tessellations around gas particles to fill the entire volume of the simulation. These tessellations are adaptive to resolve a large range of spatial scales. We generate 91 particle and group snapshots between  $z = 15$  and  $z = 0$ . The group snapshots include both SUBFIND ([Springel et al. 2001](#)) and ROCKSTAR ([Behroozi et al. 2013b](#)) galaxy catalogs with merger trees created using SUBLINK ([Rodríguez-Gomez et al. 2015](#)) and CONSISTENT TREES ([Behroozi et al. 2013c](#)), respectively. For the analyses in this paper, we use the SUBFIND catalogs, which define a galaxy using 3D overdensities and satellites as galaxies that are gravitationally bound to a larger halo.

The remainder of this section describes the implementation of the baryonic physics (Sec. 2.1), cosmology (Sec. 2.2) and initial conditions (Sec. 2.3) in more detail.

### 2.1. Baryon Prescription

We employ the TNG galaxy formation model ([Pillepich et al. 2018a](#); [Weinberger et al. 2018](#)) for the two hydrodynamic simulation suites presented in this paper. This version builds upon the earlier Illustris model ([Vogelsberger et al. 2013](#); [Torrey et al. 2014](#)) by implementing a kinetic AGN feedback and a self-consistent magnetohydrodynamic model. Additionally, prescriptions for stellar formation and evolution, galactic winds, and galactic outflows have been updated from the Illustris model. The TNG model has been tested across a wide range of environments and resolutions and reproduces several observational relations (e.g., [Pillepich et al. 2018b](#); [Springel et al. 2018](#); [Naiman et al. 2018](#); [Nelson et al.](#)

[2019](#); [Marinacci et al. 2019](#)). This subsection briefly reviews the relevant TNG prescriptions for this study—see [Pillepich et al. \(2018a\)](#) for a detailed discussion of the full TNG model.

Feedback prescriptions are one of the significant systematic uncertainties in hydrodynamical codes when modeling DM halo properties. In dwarf systems, for example, stellar feedback can redistribute the DM in the central regions of the galaxy, creating cored as opposed to cuspy density profiles depending on the mass of the host (e.g., [Lazar et al. 2020](#)). The suites presented here adopt the same fiducial parameters that define the TNG model, except for three that specifically affect SN and AGN feedback. We vary parameters that adjust the strength of feedback over a wide range to both bracket uncertainties when analyzing a particular DM model and to avoid imposing restrictive priors on machine-learning models.

The TNG subgrid model for SN feedback accounts for the ejection of winds, which move gas throughout a galaxy. There are two key inputs to this model: the mass-loading factor,  $\eta_w$ , and the stellar wind velocity,  $v_w$ . The former, which is directly related to the wind energy, is the ratio with which gas mass is lost through winds versus direct star formation. Updated from Illustris, the TNG model allows for a subdominant portion of the available energy from stellar feedback to be thermally released. Specifically, the mass-loading factor is given by

$$\eta_w = \frac{2}{v_w^2} e_w (1 - \tau_w), \quad (1)$$

where  $e_w$  is the specific energy available for generating winds in units of  $10^{51} \text{ erg } M_\odot^{-1}$ ,  $v_w$  is the wind velocity (defined below), and  $\tau_w$  is the fraction of energy that is released thermally. While  $\tau_w$  is fixed to its fiducial value of 0.1, we vary the specific energy logarithmically between  $e_w \in [0.9, 14.4]$ . The fiducial value for  $e_w$  in the TNG model is 3.6.

Increasing the strength of SN winds results in a decrease in the cosmic star formation rate density (SFRD) for  $z \gtrsim 1$  in the TNG model ([Pillepich et al. 2018a](#)). This feedback has



the greatest effect on low-mass galaxies with  $M_* > 10^{10} M_\odot$ , which comprise a significant portion of the SFRD for  $z \gtrsim 2$  (Torrey et al. 2014). At later times, AGN feedback can dominate changes to the cosmic SFRD by suppressing star formation in massive halos (Torrey et al. 2014). For a simulation of the same size and resolution as those presented in Section 4, a  $2\times$  increase in  $e_w$  results in a lower SFRD at early times that is nearly consistent with observations (Pillepich et al. 2018a). The same increase to  $e_w$  also reduces the stellar mass content of MW-mass ( $M_{\text{halo}} \sim 10^{12} M_\odot$ ) below the observed values (Pillepich et al. 2018a). Similar parameter variations within the Illustris model to those that we adopt here result in SFRDs that are nearly consistent with observations (Vogelsberger et al. 2013, 2014a). Within the TNG model, simultaneous variations of a  $2\times$  increase in  $e_w$  and an increased  $v_w$  to reproduce the fiducial value of  $\eta_w$  results in a SFRD that is too low across all redshifts (Pillepich et al. 2018a). We choose to extend the parameter space of  $e_w$  to  $4\times$  the fiducial value so that extreme variations in this one parameter will cause the simulations to be inconsistent with observations and simultaneous variations with our other SN variable, discussed below, can result in the fiducial mass loading factor.

The speed of the galactic winds created by SN feedback is scaled by the local DM velocity dispersion,  $\sigma_{\text{DM}}$ , in a redshift-dependent manner:

$$v_w = \max \left[ \kappa_w \sigma_{\text{DM}} \left( \frac{H_0}{H(z)} \right)^{1/3}, v_{w,\text{min}} \right], \quad (2)$$

where  $\kappa_w$  is a dimensionless normalization factor,  $H(z)$  is the Hubble parameter, and  $v_{w,\text{min}}$  is the minimum wind speed.  $\sigma_{\text{DM}}$  is computed from the 64 nearest DM particles to the SN and  $v_{w,\text{min}} = 350$  km/s. We vary the dimensionless normalization factor logarithmically in the range  $\kappa_w \in [3.7, 14.8]$ . The fiducial value for  $\kappa_w$  in the TNG model is 7.4.

Increasing the speed of the SN winds,  $\kappa_w$ , results in a decrease in the cosmic SFRD for  $z \lesssim 4$  where the most extreme variation is inconsistent with observations below  $z \sim 3$  (Pillepich et al. 2018a). Similar to a  $2\times$  increase in  $e_w$ , increasing  $\kappa_w$  by  $2\times$  results in a decrease in the stellar mass content of MW-mass galaxies that are inconsistent with observations (Pillepich et al. 2018a). Similar parameter variations in the Illustris model to those that we adopt here result in too low SFRD at late times for  $2\times$  their fiducial value but are consistent with observations for  $0.5\times$  their fiducial value (Vogelsberger et al. 2013, 2014a). In the same spirit as the  $e_w$  parameter, we adopt a broad range for  $\kappa_w$ .

The overall strength of AGN feedback can also play an important role in shaping the DM halos around galaxies (Anbajagane et al. 2022). In the TNG model, a black hole with a mass of  $8 \times 10^5 h^{-1} M_\odot$  is seeded at the center of a halo that reaches a mass of  $5 \times 10^{10} h^{-1} M_\odot$ . Accretion onto the black

hole is modeled by a spherical Bondi-Hoyle model (Bondi & Hoyle 1944) and feedback is implemented for two phenomenologically different states. The low-accretion state corresponds to a system where hot gas accretes isotropically onto the black hole (Shapiro et al. 1976; Ichimaru 1977), while the high-accretion state corresponds to cooler gas that accretes through a disk (Shakura & Sunyaev 1973). The high-accretion state dominates for galaxies with stellar mass  $\lesssim 10^{10.5} M_\odot$ , while the low-accretion state takes over at higher stellar masses (Weinberger et al. 2017). Therefore, for the MW zoom-in suite, the high-accretion state is the more relevant of the two. This is also true for the uniform-box suite, where the number of halos with stellar mass greater than  $\sim 10^{10.5} M_\odot$  is typically  $\sim 40$  and the number of halos less than  $\sim 10^{10.5} M_\odot$  with a black hole is typically  $\sim 10^3$  per simulation. For this reason, we only vary the high-accretion state here.

In this state, the feedback energy released from accretion is given by

$$\Delta \dot{E} = \epsilon_{f,\text{high}} \epsilon_r \dot{M}_{\text{BH}} c^2, \quad (3)$$

where  $\epsilon_{f,\text{high}}$  is the fraction of energy that is transferred to the nearby gas,  $\epsilon_r$  is the radiative efficiency, and  $\dot{M}_{\text{BH}}$  is the black hole accretion rate. All parameters are kept fixed to their TNG fiducial values, except for  $\epsilon_{f,\text{high}}$ , which is varied logarithmically in the range  $\epsilon_{f,\text{high}} \in [0.025, 0.4]$ . Its TNG fiducial value is 0.1.<sup>2</sup> We chose to vary  $\epsilon_{f,\text{high}}$  as it is a tuned parameter in the TNG model.

Increasing the strength of the AGN feedback,  $\epsilon_{f,\text{high}}$ , results in an increase in the cosmic SFRD and a corresponding increase in the stellar-mass-halo-mass relation for  $z \gtrsim 3$  (Ni et al. 2023). A  $4\times$  increase in  $\epsilon_{f,\text{high}}$  results in a  $\sim 3\times$  increase in the SFRD at late times for similar simulations to those presented in Section 4 (Ni et al. 2023). This variation is slightly larger than observational uncertainties on the cosmic SFRD (Behroozi et al. 2013a). In Weinberger et al. (2017),  $\epsilon_{f,\text{high}}$  is varied by a factor of four and shown to affect the gas fraction and star formation rate, which still remain within the bounds of observed results. The radiative efficiency, proportional to  $\epsilon_{f,\text{high}}$  in our simulations, has also been shown to depend on mass, suggesting that variations around the fiducial parameter would be needed (Davis & Laor 2011).

A number of studies have investigated how varying these astrophysical parameters can affect various galactic properties (Vogelsberger et al. 2014a; Torrey et al. 2014; Weinberger et al. 2017; Pillepich et al. 2018a; Roca-Fàbrega et al. 2021; Ni et al. 2023). We adopt parameter ranges that when varied individually are consistent with observations for the majority

<sup>2</sup> The AGN parameter we vary differs from the two parameters used in the CAMELS project, which were found to induce negligible differences across many observables (Villaescusa-Navarro et al. 2021a).

of the range that we include. The high end of our two SN parameters,  $e_w$  and  $\kappa_w$ , are shown to be inconsistent with observations when varied individually (Pillepich et al. 2018a). Only a single instance of simultaneous parameter variations is available, an increase in both  $e_w$  and  $\kappa_w$ , and is shown to be inconsistent with observations across all redshifts (Pillepich et al. 2018a). Overall, the extrema of the parameter ranges we adopt are likely to fall outside current observational uncertainties, but are still useful to mitigate prior effects.

Except for the two SN parameters and one AGN parameter described above, all other parameters in the TNG model are fixed across the simulation suites. The uniform boxes we present have a similar resolution to the TNG300-1 box, and the zoom-in simulations are at  $\sim 2\times$  lower mass resolution than TNG50-1.

## 2.2. Cosmology

For a given simulation suite, the cosmological parameters can either be fixed or varied. In cases where they are fixed, we assume a cosmology consistent with Planck Collaboration et al. (2016), where  $\Omega_m = 0.302$ ,  $\Omega_\Lambda = 0.698$ ,  $\sigma_8 = 0.839$  and  $H_0 = 100 h$  km/s such that  $h = 0.691$ . We fix  $\Omega_b$  at 0.046 for the hydrodynamic simulations and 0.0 for the N-body simulations.

In other instances, we vary  $\Omega_m$  and  $\sigma_8$  within the following ranges:  $0.1 < \Omega_m < 0.5$  and  $0.6 < \sigma_8 < 1.0$ . All other cosmological parameters are kept fixed. These ranges purposefully exceed the current uncertainties from CMB analyses (Planck Collaboration et al. 2020) so that the prior distribution has minimal impact on machine learning models. These parameter ranges are consistent with those used by the CAMELS project (Villaescusa-Navarro et al. 2021a), which has shown that machine learning models can infer cosmological values accurately across the entire range of each parameter (Villaescusa-Navarro et al. 2022b). Cosmological and astrophysical variations, when present, are varied according to a Sobol sequence (Sobol’ 1967).

## 2.3. Initial Conditions

### 2.3.1. Uniform-Box Simulations

Uniform-box simulations cover a large periodic volume at one resolution and have the advantage of including thousands to millions of galaxies across a wide range of mass and size scales. These simulations provide a cosmological context that ensures a realistic environment for galaxy formation and evolution. Moreover, the large sample size enables the study of galaxy properties that require substantial statistics to disentangle. For example, utilizing information from thousands of galaxies, one can explore whether a large sample of galactic systems can be synthesized to uncover DM models that only induce subtle changes to individual galaxy properties.

Despite these advantages, state-of-the-art uniform-box simulations are typically not able to resolve structures with a stellar mass below  $\sim 10^7 M_\odot$  (Feldmann et al. 2023; Nelson et al. 2019; Pillepich et al. 2019). While this is not necessarily a concern for investigations into large-scale structure and cosmology, it can be problematic if one wants to study DM models where deviations from CDM occur on sub-galactic scales.

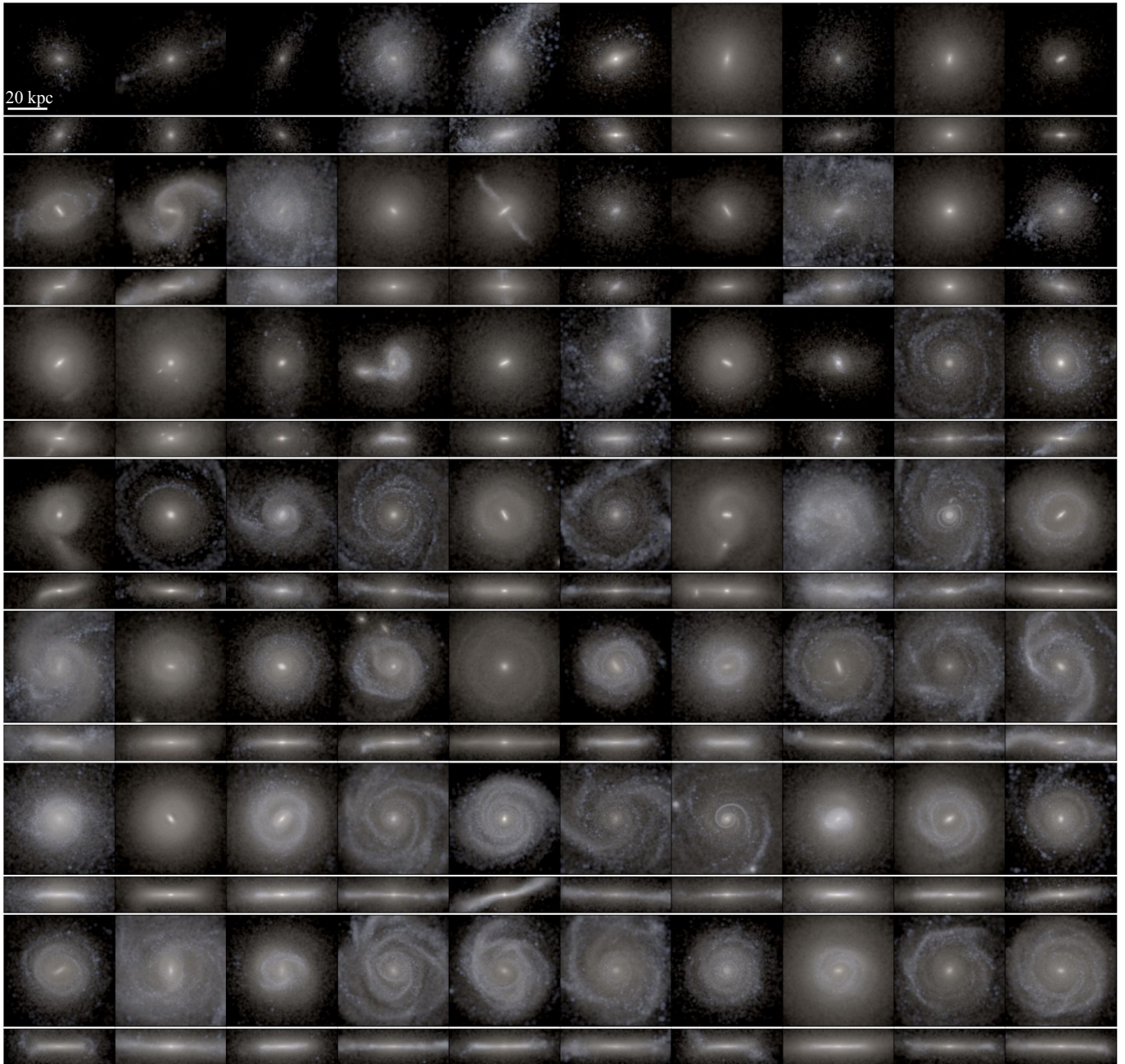
The DREAMS project does include several suites that are cosmological boxes, denoted ‘Uniform Box’ in Tab. 1. Each of these simulations covers a  $(25 h^{-1} \text{ Mpc})^3$  volume with periodic boundary conditions and starts from a  $2 \times 256^3$  grid of DM and gas particles. The DM mass resolution for these simulations is  $7.81 \times (\Omega_m/0.302) \times 10^7 h^{-1} M_\odot$ . The baryon mass resolution, corresponding to the average mass of the star particles throughout the simulation, is  $1.27 \times 10^7 h^{-1} M_\odot$ . The spatial resolution is determined by the gravitational softening factor that is applied to reduce numerical noise. The uniform-box simulations have a spatial resolution of  $1.0 h^{-1} \text{ kpc}$  at redshift  $z = 0$ .

The initial conditions for each simulation are created with NGGENIC (Springel et al. 2019). NGGENIC begins with a glass distribution of particles that evolves according to the equations for second-order Lagrangian perturbation theory up to  $z = 127$ . The glass distribution is convolved with Gaussian random noise to create a unique distribution of matter from which to start the simulations. Each simulation in the DREAMS project is given a different random seed to create the Gaussian random noise field ensuring that each simulation starts from a unique distribution of DM perturbations.

### 2.3.2. Zoom-in Simulations

We also introduce simulations that target MW-mass galaxies at much higher resolution. The advantage of these so-called “zoom-in” simulations is that they can better resolve the inner structure and dynamics of a small sample of galaxies for a reduced computational cost. This more detailed view of the internal dynamics of a MW system is imperative for testing deviations from CDM.

For our first suite of zoom-in simulations, we focus on MW-mass galaxies. Currently, the largest sample of high-resolution MW-mass galaxies is the TNG50 simulation (Nelson et al. 2019; Pillepich et al. 2019). The TNG50 simulation contains 198 MW-mass galaxies with a DM mass resolution of  $6.5 \times 10^5 h^{-1} M_\odot$  and a baryon resolution of  $1.2 \times 10^5 h^{-1} M_\odot$ . The zoom-in suites in the DREAMS project include  $\sim 5\times$  as many MW-mass galaxies at  $2\times$  lower resolution while also varying the DM and astrophysical prescriptions between the simulations. These are not direct MW-analogs as they are chosen to not be part of a local group that includes a nearby massive neighbor (for computational reasons). Future work



**Figure 1.** Three-dimensional stellar light reconstructions of the central galaxies from a small sample ( $\sim 7\%$ ) of the WDM zoom-in suite that includes variations to the DM particle mass and the TNG galaxy formation model. The images are ordered (from top left to bottom right) by increasing disk-to-stellar-mass ratios and consist of a wide range of systems ranging from red spherical to blue disk-dominated galaxies. The images are made from the  $i$ ,  $r$ ,  $g$  Sloan broadband filters. Each image shows a  $50 \times 50 \times 15 \ h^{-3} \text{ kpc}^3$  volume. Each volume is projected along the short axis for the face-on images or along one of the long axes for the edge-on images.

will explore possibilities of extending the suite to include additional environments with analogs of Andromeda, M31.

We use MUSIC (Hahn & Abel 2011) to generate the initial conditions for the zoom-in simulations in the suite. MUSIC works similarly to NGGENIC in that it first convolves a uniform particle distribution with a Gaussian random field and then uses second-order Lagrangian perturbation equations to evolve the particles to  $z = 127$ . Each MW-mass target is

chosen at random from a different  $100 \ h^{-1} \text{ Mpc}$  box with an isolation criterion of no neighbors more massive than  $7.2 \times 10^{11} \ h^{-1} \text{ M}_\odot$  within  $1 \ h^{-1} \text{ Mpc}$ . The high-resolution regions of the zoom-in simulations presented here have a DM mass resolution of  $1.2 \times 10^6 \ h^{-1} \text{ M}_\odot$  and a spatial resolution of  $0.31 \ h^{-1} \text{ kpc}$  at  $z = 0$ . The average baryon mass resolution is  $1.9 \times 10^5 \ h^{-1} \text{ M}_\odot$ . Further details on the multi-step proce-



dures required to create the initial conditions for the zoom-in simulations are provided in Appendix A.

### 3. EMULATING WDM SATELLITE COUNTS

This section introduces the zoom-in suite focused on WDM MW-mass galaxies. This is the largest sample (to date) of MW-mass galaxies that have been simulated with a realistic galaxy formation model at high resolution. WDM differs from CDM by including free-streaming velocities that suppress early-universe density perturbations at small scales because the DM escapes from potential wells (Bode et al. 2001). We consider thermal-relic WDM where the initial velocities correspond to the time at which the particles decouple from the surrounding plasma. Although non-thermal production mechanisms for WDM are also possible (Banerjee et al. 2023), the thermal scenario provides a convenient starting point for its simplicity: the only relevant DM model parameter is the fundamental WDM particle mass.

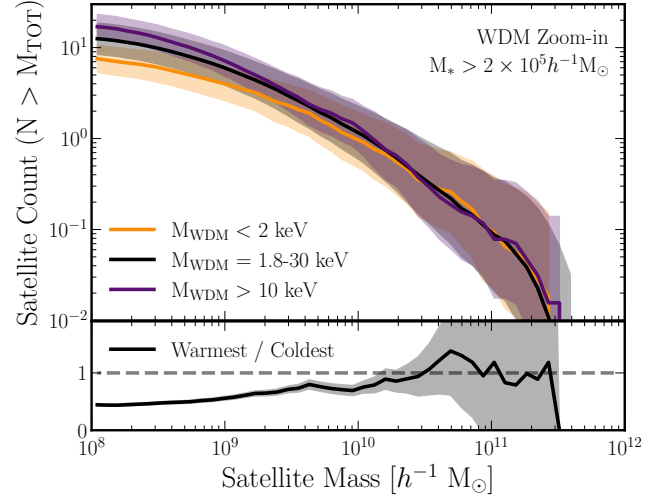
#### 3.1. WDM Zoom-In Suite

The WDM zoom-in suite consists of 1024 unique MW-mass galaxies at a baryon (DM) mass resolution of  $1.9 \times 10^5$  ( $1.2 \times 10^6$ )  $h^{-1} M_{\odot}$ .<sup>3</sup> In generating this suite, the baryonic parameters (two SN and one AGN) are varied, but the cosmological parameters are kept fixed. Additionally, the WDM particle mass is varied uniformly in  $M_{\text{WDM}}^{-1}$  between 0.033 and  $0.555 \text{ keV}^{-1}$  (1.8–30 keV). For an overview of the WDM implementation in DREAMS and a discussion of the impact of artificial clumping, see Appendix B.

For the galaxies in the suite, the central halo mass has a mean and standard deviation of  $(8.8 \pm 1.2) \times 10^{11} h^{-1} M_{\odot}$ . The host disk properties exhibit considerable variation in morphology due to the spread in the baryonic physics parameters. For example, the stellar mass of the disk<sup>4</sup> has a mean and standard deviation of  $(4.1 \pm 2.8) \times 10^{10} h^{-1} M_{\odot}$ . Additionally, one can consider the disk-to-total-mass ratio (D/T), which quantifies the ratio of co-rotating to total stellar mass within 50 kpc of the host’s center. Across all simulations in the suite, the central galaxies have an average D/T and 16–84% spread of  $0.43^{+0.20}_{-0.22}$ . This average aligns well with other simulated MW-mass halos, but with a larger range likely due to the parameter variations and large sample size used here (Obreja et al. 2016; Grand et al. 2017). Roughly 40% of the galaxies have a disk-to-total-mass ratio greater than 0.5. Current esti-

<sup>3</sup>This value is different than the one quoted in Tab. 1 because  $\Omega_m$  is now fixed at 0.301712.

<sup>4</sup>To calculate the disk mass, we first calculate the angular momentum about the system’s eigenvector of the moment of inertia tensor that aligns with the principle angular momentum axis for each star particle within 50 kpc of the host. Any material rotating with at least 70% of the maximum angular momentum is designated as being part of the disk.



**Figure 2.** The satellite mass function for galaxies in the zoom-in WDM suite, which is discussed in Section 3. We include all satellites with  $M_{\text{halo}} > 10^8 h^{-1} M_{\odot}$  and  $M_* > 2 \times 10^5 h^{-1} M_{\odot}$ , which are within 300 kpc of the central Milky Way-mass galaxy. The data is subdivided into two additional subgroups according to WDM particle mass:  $M_{\text{WDM}} < 2 \text{ keV}$  (orange) and  $M_{\text{WDM}} > 10 \text{ keV}$  (purple). Each subset contains  $\sim 100$  simulations. For each case, the average counts are shown by the solid line and the average  $1\sigma$  deviation is shown as the shaded band. The bottom panel shows the ratio between the low-WDM-mass (orange) and high-WDM-mass (purple) subsets. The spread in the ratio is indicated by the shaded band and becomes large above  $\sim 2 \times 10^{10} M_{\odot}$  due to the small number of satellites in this mass range.

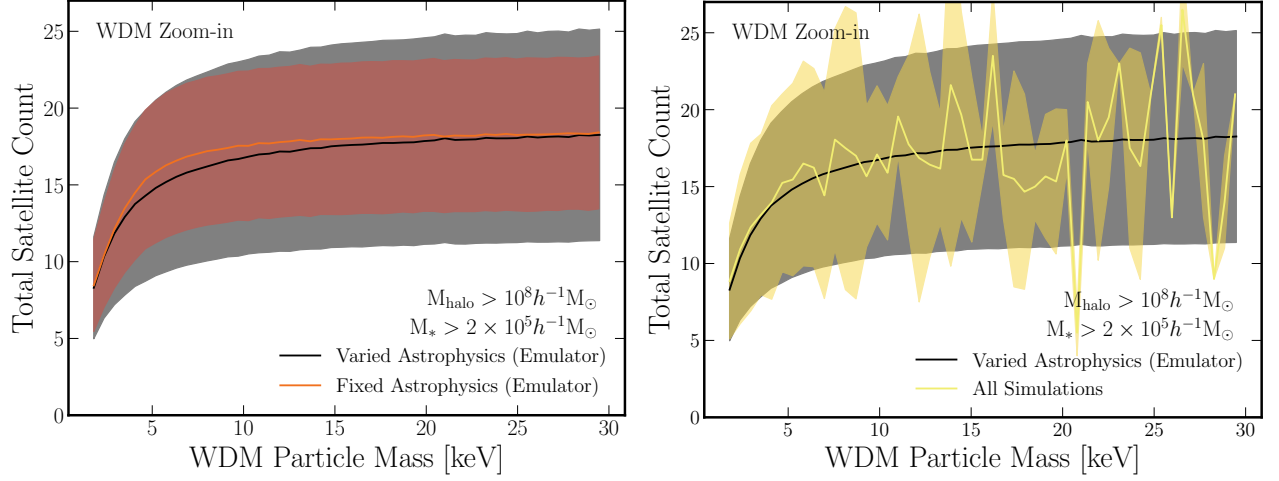
mates of the MW’s D/T ratio are near 0.8, toward the higher end of ratios exhibited in this suite (Cautun et al. 2020).

Figure 1 presents a small sample ( $\sim 7\%$ ) of MW-mass galaxies from the WDM zoom-in suite depicting face-on and edge-on views. The galaxies are rotated so that the angular momentum, computed using all star particles within 10 kpc of the potential minimum, points toward the viewer. The images are then ordered (from top left to bottom right) by increasing D/T ratio. The three-color image is comprised of the Sloan  $g$ ,  $r$ , and  $i$  filters as the blue, green, and red components of the image, respectively, with no correction for dust extinction. Overall, the population of MW galaxies in this suite covers a wide range of morphologies from red spherical galaxies to blue disk-dominated systems.

#### 3.2. Emulator Results

As a first application of the WDM zoom-in suite, this section explores how the number of satellites around a MW-mass halo depends on the WDM particle mass, as well as cosmic variance and other astrophysical uncertainties. This investigation includes all satellites that have a total DM halo mass of at least  $10^8 h^{-1} M_{\odot}$  and a stellar mass of at least  $2 \times 10^5 h^{-1} M_{\odot}$ . All satellites must also be within 300 kpc of the central galaxy. Since star formation is inefficient in





**Figure 3.** Emulator results from the neural network trained on the MW zoom-in simulation suite with varied astrophysics and WDM particle mass. *Left:* The number of satellites with  $M_{\text{halo}} > 10^8 h^{-1} M_{\odot}$  and  $M_{*} > 2 \times 10^5 h^{-1} M_{\odot}$  around a MW-mass galaxy as a function of WDM particle mass for the fiducial astrophysical parameters (orange) and varied astrophysical parameters (black). The solid line represents the mean value across all galaxies in the sample with the shaded region covering the  $\pm 1\sigma$  spread. The uncertainty on the fixed-astrophysics results (orange) represents the expected spread from cosmic variance. The uncertainty on the varied-astrophysics results (black) represents the combined uncertainty from cosmic variance and uncertainty from SN and AGN feedback modeling. *Right:* The total number of satellites around a MW-mass galaxy predicted from the simulations (orange) and emulator (black) with the same satellite selection as the left panel. Simulations are grouped in 0.5 keV bins including all astrophysics variations. Emulator results for both panels are sampled at 0.05 keV intervals and then smoothed with a top-hat kernel with a width of 0.5 keV to match the bin size of the simulation results.

low-mass halos (compared to  $10^{10}-10^{13} h^{-1} M_{\odot}$  halos), most galaxies in the suite host  $\mathcal{O}(10)$  dark satellites that are not included in this analysis.

The top panel of Fig. 2 shows the satellite mass function averaged over all MW-mass galaxies in the zoom-in suite (black). We also split the dataset into two subsets that include the  $\sim 100$  warmest ( $M_{\text{WDM}} < 2$  keV; orange) and coldest ( $M_{\text{WDM}} > 10$  keV; purple) WDM models in the suite. The solid line corresponds to the average satellite mass function from each subset and the bands correspond to the  $\pm 1\sigma$  spread, encompassing variations in host merger history, the WDM particle mass, and the three free astrophysical parameters. The bottom panel of Fig. 2 shows the ratio of warmest (orange) to coldest (purple) subsets where the solid line shows the ratio of the averages from the top panel and the shaded region represents the  $\pm 1\sigma$  spread about the average.

Lower WDM particle masses lead to a clear effect on the number of satellites around a MW-mass galaxy. For  $\sim 10^{10} M_{\odot}$  galaxies, there is a  $\sim 30\%$  suppression from the coldest to warmest subgroups. For less massive satellites, down to  $\sim 10^8 M_{\odot}$ , there is a  $\sim 50\%$  reduction in the number of satellites. Even accounting for variance from the spread in astrophysical parameters, there is a clear separation between the warmest and coldest datasets for galaxies less massive than  $\sim 10^{10} h^{-1} M_{\odot}$ , as is expected for these WDM masses (Lovell et al. 2014; Read et al. 2017).

In total, the suite of WDM zoom-ins contains  $\sim 15,200$  luminous satellite galaxies and this large dataset can be used as a

training set for a variety of machine learning applications. For this example, we build a neural network emulator that acts as a multi-dimensional interpolator, taking as input the simulation parameters and outputting the number of satellites around a MW-mass halo. This method can be used to explore how a variety of astrophysical features, such as star formation rates, bar sizes, satellite mass functions, etc. change with each of the four parameters varied in this simulation suite. While limited to interpolating within the parameter range provided, these emulators provide a useful tool for understanding complex and multidimensional trends within the data. See Appendix C.1 for details on the architecture, training, and validation.

Figure 3 shows the number of luminous satellites with  $M_{\text{halo}} > 10^8 h^{-1} M_{\odot}$  and  $M_{*} > 2 \times 10^5 h^{-1} M_{\odot}$  around the MW-mass host as a function of the WDM particle mass in keV. In the left panel, emulator results are shown for the number of satellites when only the WDM particle mass is varied (orange) and when all three astrophysical parameters ( $\epsilon_w$ ,  $\kappa_w$ , and  $\epsilon_{f,\text{high}}$ ) are also varied (black). For the fiducial astrophysics results, the emulator is sampled every 0.05 keV across the range of WDM particle masses. The results are then smoothed with a top-hat filter with a kernel width of 0.5 keV. For the varied astrophysical results, the emulator is sampled  $10^3$  times every 0.05 keV where each of the 100 samples has a unique set of astrophysical parameters. The results for the varied astrophysics model are then smoothed with a top-hat filter with a kernel width of 0.5 keV.

In Fig. 3, each solid line corresponds to the average number of satellites at a given WDM mass, and the shaded region covers the  $\pm 1\sigma$  spread in the data. For the “Fixed Astrophysics” results, the spread corresponds to cosmic variance. At 30 keV, the coldest model, the emulator predicts an average of  $16.9 \pm 4.9$  satellites around the host. Decreasing the WDM particle mass has little effect on the number of satellites down to  $M_{\text{WDM}} \sim 10$  keV. Below this point, the number of satellites diminishes quickly. At  $M_{\text{WDM}} \sim 1.8$  keV, the emulator predicts an average of  $9.3 \pm 2.9$  satellites around an isolated MW-mass galaxy.

For the “Varied Astrophysics” results, the shaded band corresponds to both the cosmic variance as well as the uncertainty in the astrophysics parameters. A  $M_{\text{WDM}} = 1.8$  (30) keV model results in a satellite count of  $8.7 \pm 3.3$  ( $16.7 \pm 7.1$ ). Similar to the fiducial astrophysics results, the emulator predicts little suppression from decreasing WDM particle mass down to 10 keV.

Across the WDM particle masses investigated here, including astrophysical variations results in minimal changes to the average number of satellites predicted around an isolated MW-mass galaxy. However, including astrophysics uncertainties increases the spread in the number of satellites from an ensemble of MW-mass galaxies. The upper bound on the uncertainty band is minimally affected up to WDM masses of 6 keV, after which variations on the SN and AGN feedback parameters increase the spread. The lower bound on the uncertainty band increases faster, after WDM masses of  $\sim 2.5$  keV.

In the right panel, the emulator results for the “Varied Astrophysics” dataset are shown again (black), along with the binned simulation results (orange). The simulation results are binned in 0.5 keV bins with the same selection criteria as the emulator results. The average and spread in the number of satellites agree well between the simulation and emulator results at low WDM mass. At high WDM mass, the number of simulations per bin decreases due to the uniform sampling in  $M_{\text{WDM}}^{-1}$ . Within this range, the simulation results become more stochastic, but the emulator results show a smooth relation across all WDM particle masses.

Investigating how each astrophysical parameter affects the number of satellites around a MW-mass host reveals that  $\epsilon_w$ , which is related to the SN wind energy, has the largest impact on the number of satellites across all WDM particle masses. Across the range of  $\epsilon_w \in [0.9, 14.4]$ , the average number of satellites for a 30 keV WDM model ranges from 28 (low  $\epsilon_w$ ) to 12 (high  $\epsilon_w$ ) satellites. The parameter with the second-most significant impact is  $\kappa_w$ , which is related to the SN wind velocity. Across the range of  $\kappa_w \in [3.7, 14.8]$ , the average number of satellites for a 30 keV WDM model ranges from 15 (low  $\kappa_w$ ) to 20 (high  $\kappa_w$ ). The parameter that varies the

AGN feedback,  $\epsilon_{f,\text{high}}$ , has essentially no effect across the range that we vary.

Equation 1 defines the mass loading factor, which relates  $\epsilon_w$  and  $\kappa_w$  and determines how much mass each SN will expel from a satellite galaxy. Based on this equation, one might expect the wind velocities to have a greater impact on the number of satellites than the wind energy, since  $\eta$  scales as  $\epsilon_w$  versus  $\kappa_w^{-2}$ . This intuition is inconsistent with the emulator results, however. One possible explanation is that the wind velocities have a minimum allowed value  $v_{w,\text{min}}$ . For low-mass systems, the wind velocity may be capped at this value, even given the variations on  $\kappa_w$ . Additionally, at higher wind speeds, the gas ejected from the SN has less time to interact with surrounding material, which can further dampen the effect.

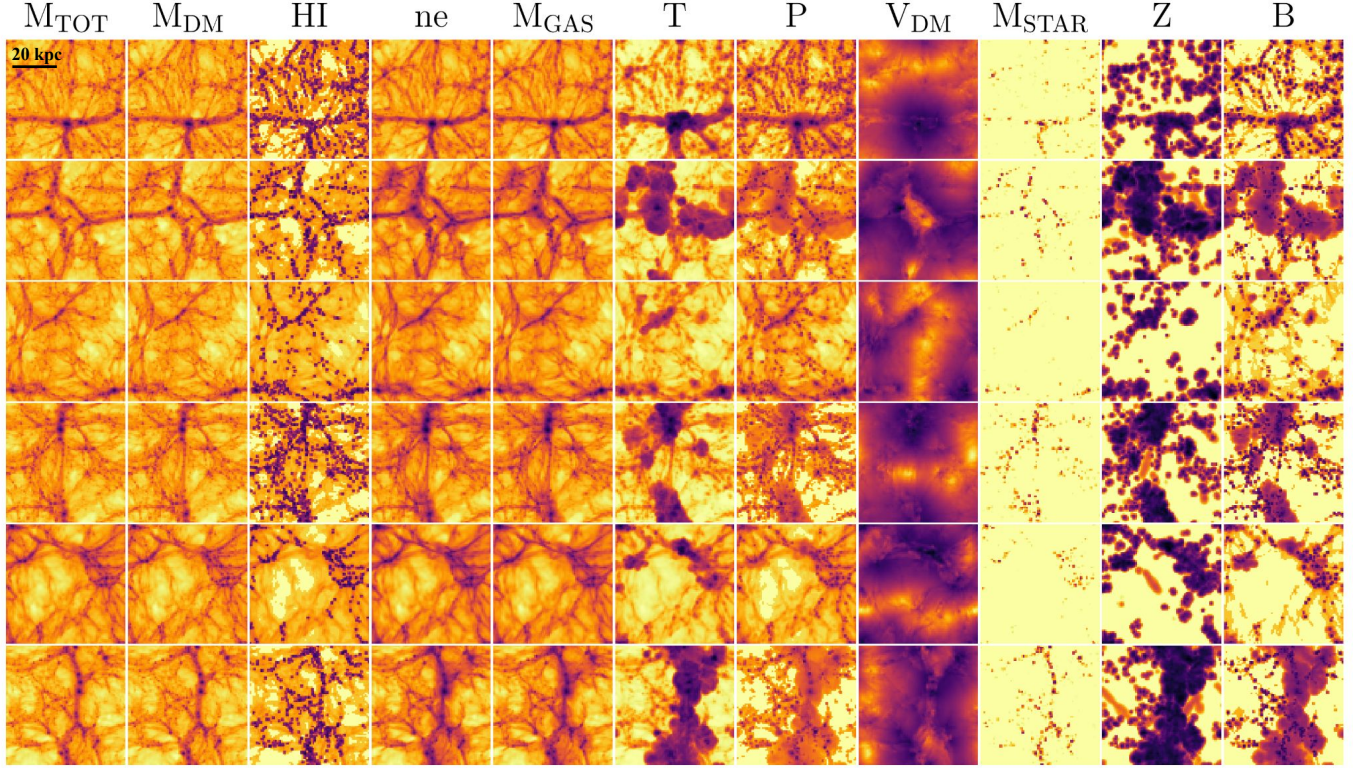
Unsurprisingly, the variations in AGN feedback do not affect the number of satellites around a MW-mass galaxy because nearly all the satellites are not massive enough to support a black hole that produces significant feedback. Similarly, the super-massive black hole in the central galaxy does not produce enough feedback to regularly disrupt surrounding satellites. While it is possible that a satellite crossing through the jet of an AGN could be disrupted, this does not appear to be a significant factor around the MW-mass systems studied here.

#### 4. PREDICTING WDM MASS FROM GALAXIES

This section explores the WDM uniform-box TNG suite in more detail. It consists of 1024 hydrodynamic cosmological simulations, each with different cosmology, astrophysics, initial density field, and WDM particle mass. Each row of Fig. 4 shows a  $25 \times 25 \times 5 h^{-3} \text{ Mpc}^3$  slice from five random simulations at  $z = 0$ . The columns correspond to a sample of properties that are tracked for each simulation, which are described in the figure caption. See Appendix C.2 for a description of how these images are created and a more detailed description of the different fields displayed in Fig. 4.

So far, constraints on the WDM particle mass at  $z = 0$  have largely come from observations of satellite properties (Nadler et al. 2021b), with other constraints coming from higher redshifts through the Ly $\alpha$  forest (Iršič et al. 2017), strong lensing (Gilman et al. 2020; Hsueh et al. 2020), and UV luminosity functions (Corasaniti et al. 2017; Sabti et al. 2022). Each of these constraints relies on various summary statistics believed to capture information that constrains the WDM particle mass. Recent investigations have produced better constraints by combining multiple statistics into one joint constraint (Nadler et al. 2021a; Enzi et al. 2021).

This section considers the constraining power of field-level data that is based on the DREAMS uniform-box suite. Specifically, the data consists of 2D density projections of a  $25 \times 25 \times 5 h^{-3} \text{ Mpc}^3$  volume and a mass resolution of



**Figure 4.** Image slices from the WDM uniform-box suite with varied cosmology and TNG physics. Each column represents a three-dimensional reconstruction of the given property that is then projected onto two dimensions: total matter density,  $M_{\text{TOT}}$ ; dark matter density,  $M_{\text{DM}}$ ; neutral hydrogen gas density, HI; electron number density, ne; baryonic gas density,  $M_{\text{GAS}}$ ; gas temperature, T; gas pressure, P; dark matter velocity modulus (speed),  $V_{\text{DM}}$ ; stellar mass density,  $M_{\text{STAR}}$ ; gas metallicity, Z; and magnetic field strength, B. Each row represents a projection from a different simulation taken randomly from the suite. Each image covers a  $25 \times 25 \times 5 \, h^{-1} \text{ Mpc}$  volume projected along the short axis.

$7.81 \times (\Omega_{\text{m}}/0.302) \times 10^7 \, h^{-1} \text{ M}_{\odot}$ . The analysis does not rely on any chosen summary statistic, and the results are only limited by the amount and quality of the data given. While the example in this section is illustrative of the general technique, more sophisticated iterations in the future can generate mock data catalogs based on the DREAMS suites to make concrete predictions for any observable of interest. This will enable one to understand the constraining power of a given observable in testing a particular DM model while marginalizing over uncertainties from astrophysics and cosmology.

#### 4.1. General WDM Features Across Simulations

To build intuition for the content of the WDM uniform-box suite, we begin with a general discussion of the simulated galaxies. Figure 5 shows the halo mass function (HMF) for all halos more massive than  $10^{10} \, h^{-1} \text{ M}_{\odot}$ , the lowest-mass resolved in these simulations, with a stellar mass of at least  $5 \times 10^6 \, h^{-1} \text{ M}_{\odot}$ . The turn-over at  $10^{11} \, h^{-1} \text{ M}_{\odot}$  arises because lower-mass halos have a greater chance of not meeting the stellar mass cut that is applied. As in Fig. 2, the results are provided for all WDM masses (black) as well as the coldest subset with  $M_{\text{WDM}} > 10 \text{ keV}$  (purple) and the warmest subset with  $M_{\text{WDM}} < 2 \text{ keV}$  (orange). Solid lines denote the average halo counts, whereas the bands denote the  $\pm 1\sigma$  spread arising

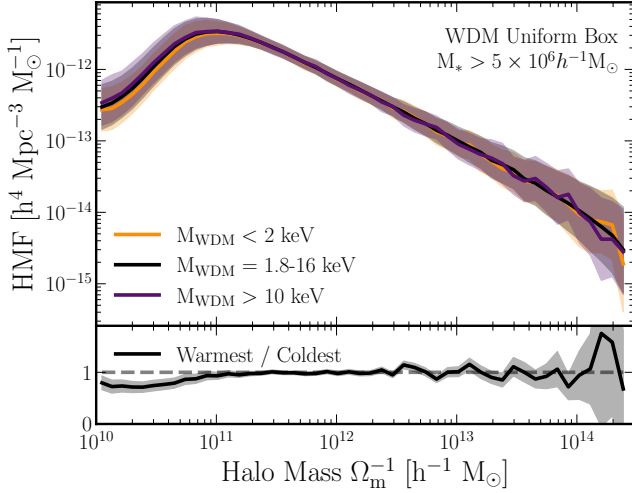
from cosmic variance as well as uncertainties in astrophysics and cosmological parameters.

As expected, the lowest-mass halos near  $10^{10} \, h^{-1} \text{ M}_{\odot}$  have the strongest suppression ( $\sim 30\%$ ), on average, between the warmest and coldest models. Any signal for the highest-mass galaxies is dominated by cosmic variance as these simulations lack a large number of groups and clusters. However, WDM is not expected to impact halo abundance at these scales.<sup>5</sup>

Clearly, the imprint of WDM mass on the halos resolved within the uniform-box simulations is rather weak. We note that the simulation-to-simulation scatter may be larger than current uncertainties on baryonic feedback due to the large range over which the parameters are varied. However, the WDM signal at the mass scales probed in these simulations only imprints a  $\sim 30\%$  reduction in the halo mass function when averaged over 100 simulations. Since the WDM signal

<sup>5</sup> We have also compared the stellar-mass-to-halo-mass (SMHM) relation for all host halos more massive than  $10^{10} \, h^{-1} \text{ M}_{\odot}$  that have a stellar component, as well as the cosmic star formation rate density (SFRD). In both of these cases, there are no significant deviations between the warmest and coldest subgroups that cannot be accounted for by cosmic variance or differences in the astrophysics and cosmological parameters.





**Figure 5.** The halo mass function (HMF) for the uniform-box WDM suite presented in Section 4. We include all halos with  $M_{\text{halo}} > 10^{10} h^{-1} M_{\odot}$  and  $M_{*} > 5 \times 10^6 h^{-1} M_{\odot}$ . The data is also subdivided into two subgroups according to WDM particle mass:  $M_{\text{WDM}} < 2$  keV (orange) and  $M_{\text{WDM}} > 10$  keV (purple). Each subset contains  $\sim 100$  simulations. For each case, the average counts are shown as solid lines and the average  $\pm 1\sigma$  deviation is shown as the shaded band. The bottom panel shows the ratio between the low-WDM-mass (orange) and high-WDM-mass (purple) subgroups. The spread in the ratio is large above  $\sim 10^{13} M_{\odot}$  due to the small sample of halos in this mass range.

is so weak, these simple statistics on their own cannot be used to infer WDM particle masses.

However, more information about the WDM is present in these simulations than might otherwise be captured by the statistics used in e.g., Fig. 5, and advanced analysis methods may be more successful at extracting the signal. To illustrate this point, consider how much WDM “signal” is present in a given area within the simulation. Figure 6 compares two simulations with the same initial density field, cosmology, and astrophysics, but different WDM mass (1.8 versus 10 keV).<sup>6</sup> The left and middle panels show the 2D projection of the matter density field from each simulation. For each image, a section from the upper left corner is expanded to show the ratio of this image with the same one taken from a CDM simulation. The ratio is displayed on a pixel-by-pixel basis with a diverging color map where less (more) dense regions in the WDM projections relative to CDM are colored red (blue). Each pixel represents a  $10 \times 10 h^{-2} \text{ kpc}^2$  area with a depth of  $5 h^{-1} \text{ Mpc}$  projected into two dimensions. The dynamic range of the color map is matched for both the  $M_{\text{WDM}} = 1.8$  and 10 keV case such that the strongest colors align with a  $5\times$

<sup>6</sup>These two examples are not taken from the DREAMS suite and are provided to illustrate differences when the initial conditions and baryonic/cosmological parameters are fixed.

deviation in either suppression or elevation. The right-most panel of Fig. 6 shows a histogram of the pixel ratios from the entire image for each WDM simulation.

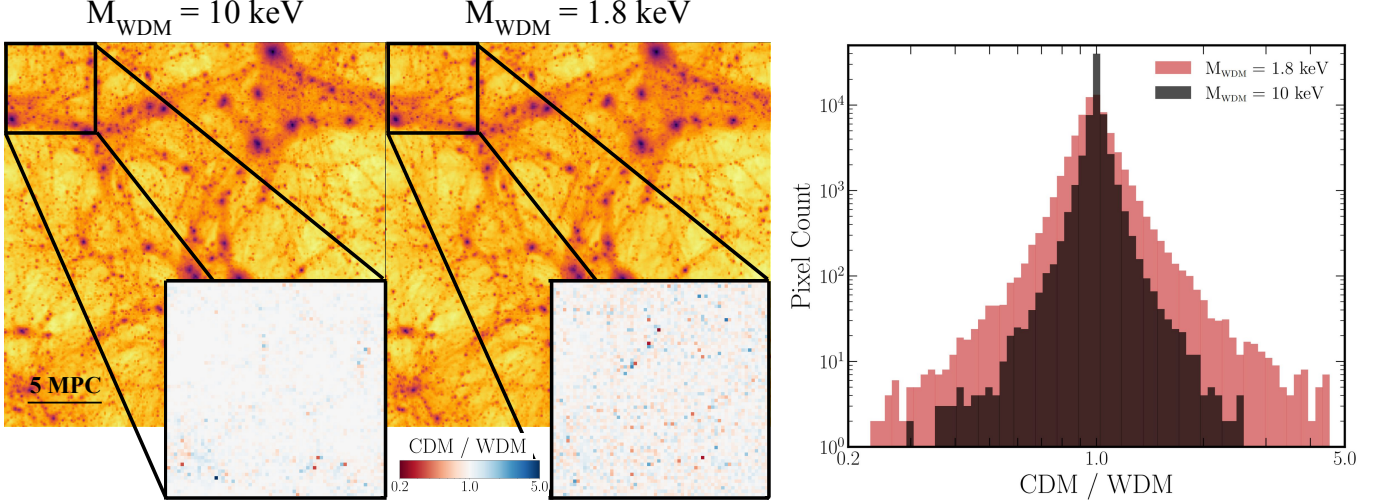
Overall, there is less deviation from CDM for the  $M_{\text{WDM}} = 10$  keV model compared to the 1.8 keV case. For the 10 (1.8) keV image, 90% (66%) of the pixels are consistent with CDM to within 10%. Additionally, significant pixel deviations from CDM are more widespread for the 1.8 keV image compared to the 10 keV case, where they are localized around high-density regions. These more subtle differences between images can potentially be captured by a CNN, as described next.

#### 4.2. Inference with Convolutional Neural Networks

CNNs offer a natural way to predict DM properties from simulation data since they are designed to process 2D images within an inference framework. The ability of these networks to move from simulated to observed data is currently limited and actively researched (for a review, see Baron 2019). However, these models have been shown to successfully estimate the mass of clusters (Ntampaka et al. 2019; Ho et al. 2023), distinguish between modified gravity and  $\Lambda$ CDM cosmologies (Peel et al. 2019), infer cosmological parameters from weak-lensing data (Gupta et al. 2018; Ribli et al. 2019; Fluri et al. 2019), create a mapping from baryon to DM distributions (Zhang et al. 2019), and to locate substructure and estimate model parameters within strong lensing (Alexander et al. 2020; Diaz Rivero & Dvorkin 2020; Schuldt et al. 2021).

We adopt the architecture and training procedures from the CAMELS Multifield Dataset (CMD) (Villaescusa-Navarro et al. 2022b) to train CNNs on field-level astronomical data. This architecture was used in Villaescusa-Navarro et al. (2021c,b) and found to perform well at predicting cosmological parameters from astrophysical images. In Rose et al. (2024), this architecture was used to predict the WDM particle mass from similar images to those used here, except taken from N-body simulations where the WDM particle mass,  $\Omega_{\text{m}}$ , and  $\sigma_8$  were varied simultaneously. We note that the images we analyze are created from full field-level data so there are no selection criteria imposed on the data. Hence, this dataset includes low-mass halos, but we do not expect unresolved features of these halos to affect our results given the pixel resolution of the images, Appendix C.2 provides a brief summary of the key features of the architecture and image creation; the interested reader can go to Villaescusa-Navarro et al. (2022b) for a more detailed description.

We first train and test the CNN to predict the WDM particle mass from images of the total matter density field. The results of this analysis are shown in the top panel of Fig. 7. Each point on the plot shows the predicted mean of the posterior distribution from one image from the test dataset. The errorbars show the  $1\sigma$  uncertainty, which includes both the



**Figure 6.** A comparison of the 2D projected total matter density for two different WDM models with  $M_{\text{WDM}} = 10$  (1.8) keV in the left and middle panel, respectively. These particular images are not taken from the DREAMS suite as they only contain differences in the WDM particle mass and do not include variations in the initial density field, baryonic feedback, or cosmology. The size of each pixel covers a  $10 \times 10 h^{-1}$  kpc area with a projected depth of 5 Mpc. The top left corner of each image is enlarged and we provide the pixel-by-pixel ratio of each with the corresponding image from a CDM simulation. The zoomed-in panels have a diverging color map where blue (red) corresponds to over (under)-dense regions. The dynamic range is fixed with a maximum variation of  $5\times$  to highlight the relative deviations from CDM. The right-most panel shows a histogram of the pixel ratios for the two WDM models. In general, the 10 keV WDM model exhibits fewer pixel-level deviations from CDM than the 1.8 keV one.

standard deviation of the posterior distribution (aleatoric uncertainty) and the uncertainty intrinsic to the machine learning method (epistemic uncertainty). The epistemic uncertainty is always subdominant to the aleatoric uncertainty. Since the simulations are sampled uniformly in  $M_{\text{WDM}}^{-1}$ , the posterior inferred by the CNN also has units of  $\text{keV}^{-1}$ . To translate the uncertainties from units of  $\text{keV}^{-1}$  to keV, we sample the uncertainty—the sum of the aleatoric and epistemic errors—as a Gaussian distribution with 1000 samples. We then take the inverse of these samples and present the  $1\sigma$  spread.

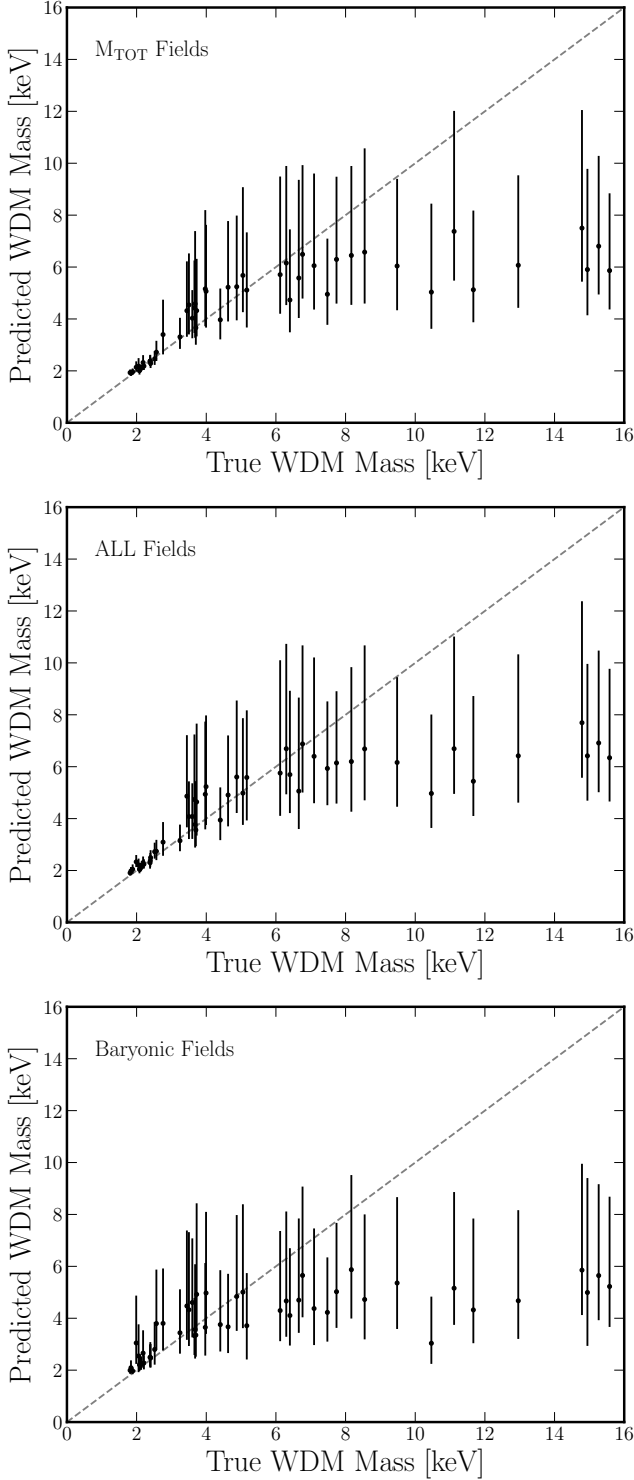
The trained neural network can predict the WDM particle mass up to 3.25 keV with less than  $\pm 1$  keV uncertainty. At greater DM particle masses, the network becomes more uncertain with some predictions being close to the true values and others plateauing at  $\sim 6$  keV. The plateau at  $\sim 6$  keV results from the network being unsure of the true particle mass above  $\sim 4$  so the network predicts the mean of the remaining distribution. This first analysis is restricted to only the total matter density field to mirror that of [Rose et al. \(2024\)](#). They found that their neural network (which utilized the same architecture used here) can make predictions with at least  $2\sigma$  confidence up to 3.5 keV with an average uncertainty of 0.6 keV. Over the same mass range, our model has a larger average uncertainty of 1.2 keV. Hence, the inclusion of a galaxy formation model (with parameter variation)

does result in an increased uncertainty on the predicted WDM particle mass.<sup>7</sup>

We can attempt to infer the WDM particle mass from more than just the total mass density field given that the IllustrisTNG model is used in the simulations. To test this, we train a CNN on a multifield dataset where each property in Fig. 4 is included as a separate channel. The results are provided in the middle panel of Fig. 7. The trained CNN predicts the WDM particle mass from the multifield dataset up to 3.25 keV with an uncertainty of less than 1 keV, matching the results when the model was trained on only the total matter density field. Hence, including the baryonic and additional DM information does not significantly improve the ability to recover the WDM mass. This result suggests that any information that is contained in the baryonic fields is degenerate with the information in the total matter density field.

To test this further, we also train a model on the baryonic fields alone. This includes all fields from Fig. 4, except  $M_{\text{TOT}}$ ,  $M_{\text{DM}}$ , and  $V_{\text{DM}}$ . These results are shown in the bottom panel of Fig. 7. In this case, the network predicts the WDM particle mass up to 2.0 keV with an uncertainty of less than 1 keV—slightly worse than the previous two examples. Therefore,

<sup>7</sup>The analysis done in [Rose et al. \(2024\)](#) was performed on images with a resolution of  $512^2$  pixels whereas this study is done on images with a  $256^2$  pixel resolution. Given that [Rose et al. \(2024\)](#) showed that pixel resolution did not have a significant effect on their results for their fixed-cosmology dataset at the simulation resolution used for our simulations, this should not have a significant effect on our results.



**Figure 7.** The results of the CNN trained to infer the WDM particle mass from fields available in the WDM uniform-box simulation, such as the total mass density (top) and the baryonic fields (bottom). The middle panel shows the results using all the fields listed in Fig. 4. Including the baryonic fields does not provide better predictive power over just the total matter density field.

while the baryonic fields do contain some information that can be used to predict the WDM particle mass, the total matter density field provides more.

To investigate which of the fields in Fig. 4 contributes the most to the predictions in Fig. 7, we train separate CNNs on each and compare the relative error of the results for WDM masses below 6.0 keV. We define the relative error as the difference between true and predicted values divided by the true value averaged over all points in the given dataset. The fields  $M_{\text{TOT}}$  and  $M_{\text{DM}}$  produce comparable results with a relative error of 0.10, the best of all the cases tested here. The next best results come from gas properties that all produce similar results, similar to those shown in the bottom panel of Fig 7. These fields (ne, HI, T, P, and  $M_{\text{GAS}}$ ) have a relative error of 0.15. The rest of the fields ( $V_{\text{DM}}$ ,  $M_{\text{STAR}}$ , Z, and B) all provide no predictive power with an average relative error of 0.95.

Overall, the most predictive power is contained in the DM density fields, which is degenerate with the information in the total matter density field. Limiting the analysis to baryonic properties does still allow one to predict the WDM particle mass, although not as well. Nevertheless, the fact that the baryonic fields do have some discriminatory power is important as they are directly linked to experimental probes. For instance, the HI field is mapped by the 21 cm line (Pritchard & Loeb 2012) and Lyman-alpha forest (Ouchi et al. 2020), the temperature field by X-rays (Schaye et al. 2000; Haardt & Madau 2012) and the thermal Sunyaev-Zeldovich effect (Mroczkowski et al. 2019; Birkinshaw 1999), and the B-field by Faraday rotation measures of radio sources (Ryu et al. 2008; Feretti et al. 2012). To definitively say which of these probes is most useful in discriminating WDM masses, one would need to translate the simulation output into mock observational data, which is left for future work.

The results presented here represent the best predictions that can be made from the data supplied to the CNN, ie. 2D density fields taken from a  $25 \times 25 \times 5 (h^{-3} \text{ Mpc}^3)$  volume with a DM mass resolution of  $7.81 \times (\Omega_{\text{m}}/0.302) \times 10^7 h^{-1} M_{\odot}$ . Understanding how predictions from more observationally-motivated summary statistics, such as the power spectrum, compare to these predictions can shed light on the extent to which a given summary statistic, or ensemble, can capture all the information present in this dataset. As shown in Rose et al. (2024), the accuracy and uncertainty of these predictions improve with increased simulation resolution. Therefore, the results that we show here represent a first step in the predictive power of this method where increasing the data size and resolution to current observational levels can help to understand whether our current observables miss important constraining information on the WDM particle mass.



## 5. EXTENSIONS TO THE DREAMS PROJECT

We have presented two case studies that utilize data from the DREAMS project to study properties of WDM while simultaneously accounting for astrophysical uncertainties. The framework illustrated through these examples consists of three modular components that can be adapted for a given problem of interest. These include the choice of (1) DM model, (2) galaxy formation model, and (3) astrophysical target. This section discusses how these separate components can be altered in future DREAMS suites.

### 5.1. Varying the Dark Matter Model

Despite decades of effort, very little is known about the particle-physics nature of DM, including how it interacts with the Standard Model or with itself. DM may be fermionic or bosonic; it can behave as a wave or as a particle; it may interact with the Standard Model via new dark forces, or only via gravity; it may be one single species, or populate a rich “dark sector”; it may be produced thermally from a hot plasma, or via non-thermal mechanisms. The possibilities are numerous and the objective of the DREAMS project is to take a strategic approach, focusing on several “simplified models” that capture key features of the different classes of theories. The goal is to build intuition for how different DM properties impact galaxy formation and evolution, possibly unveiling new key observables.

The models we plan to explore as part of the DREAMS project can be divided into two broad classes: those where the initial DM transfer function is modified and those with *in situ* effects. The example discussed in this paper falls in the first category, as WDM is the simplest model that suppresses the DM transfer function. As a natural extension, we will consider more general modifications of the transfer function within the framework of the Effective Theory of Structure Formation (ETHOS; Cyr-Racine et al. 2016; Vogelsberger et al. 2016). Within the standard ETHOS example, the dark sector consists of a DM particle and a relativistic species (e.g., dark radiation, Bøhm et al. 2002). The presence of dark radiation leads to dark-acoustic oscillations as well as DM self-interactions (van den Aarssen et al. 2012). A useful variant on this framework is the  $h_{\text{peak}} - k_{\text{peak}}$  phenomenological model proposed by Bohr et al. (2020), which isolates the early-universe ETHOS behavior by parameterizing modifications to the transfer function alone.

Separately, we will also consider simplified models that focus on the *in situ* behavior of the DM. In this direction, one important question concerns the spin-statistics of DM: is DM a fermion or a boson? The DREAMS team will explore the possibility of simulating Wave Dark Matter (i.e., ultra-light bosons), with a particular focus on Fuzzy Dark Matter (FDM) (Li et al. 2019; Dome et al. 2022; Laguë et al. 2024; Lora et al. 2012; Mocz et al. 2017; Lancaster et al. 2020; Dalal

& Kravtsov 2022), with or without self-interactions (Painter et al. 2024). Particular care will be dedicated to studying the impact of baryonic feedback for FDM solvers, building on Mocz et al. (2019), which studied for the first time the interplay between baryons and FDM using AREPO (Springel 2010; Springel et al. 2019; Weinberger et al. 2020).

As particle interactions in the Standard Model are rich and diverse, it is also plausible that new interactions arise within the dark sector. Another milestone of the DREAMS project will be the study of more complicated interactions between DM particles. This includes elastic self interactions (SIDM; Spergel & Steinhardt 2000), which have been routinely advocated as a solution to many of the small-scale “tensions” mentioned in the introduction (Tulin & Yu 2018). Even in relatively simple models with a single DM and mediator species, self-interactions of relevant strengths for galactic-scale observables can be achieved for sub-GeV DM and mediator masses (Kaplinghat et al. 2016). In general, these self interactions must be velocity-dependent to be consistent with current constraints from galaxy groups/clusters and MW satellites (Slone et al. 2021; Silverman et al. 2023). Importantly, in the region of allowed parameter space, gravothermal collapse may play a key role in SIDM halo evolution (Balberg & Shapiro 2002; Koda & Shapiro 2011; Essig et al. 2019; Nishikawa et al. 2020), affecting galaxy diversity. We will explore running hydrodynamical cosmological simulations of velocity-dependent SIDM, carefully considering numerical convergence of the halo’s evolution under gravothermal collapse (Zeng et al. 2022; Mace et al. 2024; Palubski et al. 2024).

Inelastic self interactions are also important to consider. Both exothermic and endothermic DM scattering processes have been considered in the literature (Todoroki & Medvedev 2019; Vogelsberger et al. 2019; Todoroki & Medvedev 2022; Shen et al. 2021, 2022b; O’Neil et al. 2023) and shown to affect, e.g., the central densities of galaxies. Strongly dissipative dark sectors can have even more pronounced effects, leading to unexpected structures such as dark disks and/or novel compact objects (Fan et al. 2013; Ghalsasi & McQuinn 2018; Chang et al. 2019; Ryan et al. 2022). Atomic Dark Matter (Kaplan et al. 2010; Cyr-Racine & Sigurdson 2013) provides an appealing simplified model for this case study, as it can also explain other conundrums, such as the DM-baryon coincidence (Bodas et al. 2024). Very recent works, which we plan to refine and further extend, have shown that even a small fraction of strongly dissipative DM can dramatically affect a galaxy’s baryonic distribution (Roy et al. 2023; Gemmell et al. 2023).

The simplified models outlined above illustrate promising avenues of exploration, which may be modified moving forward depending on computational feasibility (e.g. implementation strategy and/or compute time). However, even if the

specifics change, the guiding principle will be to work with a set of models that adequately capture the allowed range of DM phenomenology on galactic and sub-galactic scales.

### 5.2. Varying the Galaxy Formation Physics

This work introduces two WDM suites with AREPO that utilize the TNG model. Within this framework, we vary two parameters related to SN feedback and one related to AGN feedback. These two suites, which comprise 2048 simulations, comprise the largest number of alternative DM simulations that account for uncertainties in astrophysical parameters. However, they are limited by the (small) number of astrophysical parameters varied over, as well as the TNG physics implementation. We therefore identify two axes in which the DREAMS program can be extended: (i) expanded variations within a given subgrid model and (ii) considerations of alternative sub-grid models. This subsection details recent developments in the field along each axis and how they can be incorporated within the DREAMS project.

Recently, Ni et al. (2023) introduced a new suite of simulations that simultaneously vary 28 parameters within the TNG model, including the three considered in this work. Some example variations are: the slope of the initial mass function (IMF), the minimum stellar mass that produces a SN, and the minimum SN wind velocity. Other parameters that they do not include in their analysis, such as the time of reionization, may mimic the effects of alternative DM models (Schultz et al. 2014; Lovell et al. 2018; Shen et al. 2024) or affect the formation of low-mass galaxies (Bose et al. 2018; Dawoodbhoy et al. 2018; Katz et al. 2020). While not all simulations within Ni et al. (2023) match current observational constraints, the additional parameters were varied over ranges set to both be realistic and relevant for galaxy evolution scales. Ni et al. (2023) found that, while the additional variations resulted in a much larger range of galactic properties, their inference framework could still predict  $\sigma_8$  and  $\Omega_m$  with some success. Expanding the number of TNG parameters varied in the DREAMS suites would allow for the inclusion of e.g., uncertainties in the IMF slope or the start of reionization, which can affect small-scale observables.

However, even with these additional parameter variations, the simulations are still limited to the physics implemented within the TNG model. As one example, TNG places a temperature floor on the interstellar medium (ISM) that limits gas cooling below  $10^4$  K. This implementation does not allow for a multi-phase ISM, which can lead to more localized star formation and change how SN energy is injected into the system. In lower-mass halos, such localized feedback can expel more mass from the center of the halo, reducing the central DM density (Governato et al. 2010; Brooks et al. 2013; Lazar et al. 2020). This coring from feedback has been proposed as a mechanism to alleviate some small-scale ten-

sions, such as the diversity and too-big-to-fail problems (Di Cintio et al. 2014; Chan et al. 2015; Tollet et al. 2016; Lazar et al. 2020). Introducing suites that are simulated with models that include processes like a multiphase ISM allows one to bracket the uncertainties on small-scale structure predictions between different sub-grid implementations. One such example that we plan to explore is the Adaptive Mesh Refinement (AMR) code RAMSES (Teyssier 2002), which allows for localized stellar feedback modes.

Additionally, other codes take a different approach to solving the equations that govern galaxy formation and evolution. AREPO solves its equations of gravity using a TreePM grid (Bagla 2002). While this method provides accurate results, Shao et al. (2022) showed that a Graph Neural Network (GNN) trained on a halo catalog from one code does not always perform well when tested on a halo catalog from a different code. For example, one GNN from Shao et al. (2022) that is trained on a dataset from Gadget (Springel 2005) does not predict  $\sigma_8$  correctly when tested on a similar dataset from Enzo (Bryan et al. 2014). Similar to Arepo, Gadget also uses a TreePM grid to solve the equations of gravity. Enzo on the other hand, is an AMR code that uses a fast Fourier technique to simulate gravity (Hockney & Eastwood 1988). Other techniques used to solve gravity equations include the Fast Multipole Method (Greengard & Rokhlin 1987) implemented in PKDGrav3 (Potter et al. 2017), particle-mesh interactions implemented in CUBEP<sup>3</sup>PM (Harnois-Déraps et al. 2013), and multipole-approximations implemented in Abacus (Garrison et al. 2021). It is unclear the extent to which the subtle numerical deviations between these simulations affect the field-level results inferred from them.

Since these deviations are present in gravity calculations, they should be more pronounced once the hydrodynamic equations are included as well. The CAMELS project has tried to use machine learning to marginalize over numerical uncertainty from the galaxy formation implementation when predicting cosmological parameters (Villaescusa-Navarro et al. 2021a). They do this by training a machine learning model on output from multiple simulation codes, then evaluate robustness by testing the model on output from yet another simulation code. They concluded that adding simulation codes to the training dataset can provide more robust results than training on output from just one simulation code (de Santi et al. 2023b; Shao et al. 2023; Ni et al. 2023). Because the DREAMS suites consist of many individual simulations packaged together, we can also expand the training datasets to include simulations from multiple codes to marginalize over numerical uncertainty.

### 5.3. Varying the Zoom Target

Appendix A presents a procedure to efficiently generate large sets of zoom-in simulations, which was used to simulate

the 1024 different MW-mass galaxies presented in Section 3. Moving forward, we plan to also consider different targets for the zoom-in. While MW-mass galaxies offer many opportunities to compare to local observations, dwarfs are DM-dominated systems that are particularly sensitive to changes in DM physics (for a review, see [Simon 2019](#)). Current and upcoming surveys, such as SPARC ([Lelli et al. 2016](#)), DES ([The Dark Energy Survey Collaboration 2005](#)), ELVES ([Carlsten et al. 2022](#)), SAGA ([Geha et al. 2017](#)), Roman ([Spergel et al. 2015](#)), LSST ([LSST Science Collaboration et al. 2009](#)), ARRAKIHS ([Guzmán et al. 2022](#)), and Merian ([Luo et al. 2023](#)), offer a wealth of observations targeting dwarf systems that will benefit from theory investigations performed within the DREAMS project.

The local environment of the zoom-in galaxy is another consideration for future extensions of DREAMS. For the WDM zoom-in suite, there is an isolation criterion that removes systems with neighbors (within 1 Mpc) that are more massive than  $5 \times 10^{11} M_{\odot}$ . This requirement was put in place due to the extensive computational time (an increase of at least  $2\times$ ) needed to simulate such systems. To have more accurate MW analogs, we eventually want to simulate systems with a massive neighbor, similar to Andromeda. One promising technique is to augment a MW-mass zoom-in dataset with a smaller suite that focuses on these large-neighbor systems. Applying transfer learning methods, most of the training can be performed on the large and computationally inexpensive suite, with the machine-learned model then fine-tuned to incorporate results from the smaller and more expensive suite.

These augmented suites can also be used to extend the datasets to other more computationally intensive suites that would otherwise not be possible to execute for 1000+ simulations. Another example would be a suite of simulations at higher resolution. This would allow one to investigate newly-resolved regimes at a reasonable computational cost, such as ultra-faint satellites ([Bonnivard et al. 2015](#); [Brandt 2016](#); [Hoof et al. 2020](#)) or stellar streams ([Banik et al. 2021](#); [Pearson et al. 2022](#)).

While the WDM zoom-in suite targets MW-mass haloes with a mass of  $(8.8 \pm 1.2) \times 10^{11} h^{-1} M_{\odot}$ , an additional suite of simulations could vary this target threshold. The advantage of this would be to not rely on the MW’s halo mass estimates and to incorporate DM measurement uncertainties to match current observations. Additionally, if the mass range is extended further, the dataset could be used to compare to an ensemble of observations ranging from clusters to ultra-faint galaxies simultaneously. This suite could help to incorporate a range of observables—from the density profiles of ultra-faint galaxies to strong-lensing predictions from clusters—self-consistently to thoroughly test various DM models.

## 6. CONCLUSIONS

We introduced the DREAMS project as a new approach to exploring the impact of DM on galaxy formation and evolution. Two WDM simulation suites—one consisting of uniform boxes and the other of MW zoom-ins—were introduced. Each suite includes 1024 cosmological hydrodynamic simulations generated with the AREPO code that assumes the TNG galaxy formation model. In addition to varying over the WDM particle mass, both suites also vary over the initial density field as well as several key parameters in SN and AGN feedback. In this way, the simulations span uncertainties in both cosmic variance and astrophysics modeling. The uniform-box suite also includes variations in cosmological parameters. Notably, the DREAMS project includes the largest number of MW zoom-ins that have been generated to date, providing unprecedented statistics on the properties of classical satellites in a WDM cosmology.

We presented the results of two illustrative analyses that demonstrated the possibilities for machine-learning-focused studies. In summary:

- A neural network emulator was used to disentangle the multidimensional parameter space spanned by the simulations to quantify how the WDM mass and feedback parameters affect MW satellite counts. The analysis framework allows one to quantify how the spread in satellite counts from cosmic variance increases once uncertainties from astrophysical modeling are included. See Section 3.2.
- CNNs were used to marginalize over uncertainties in cosmology and astrophysics to predict the WDM mass from two-dimensional density projections. While the DM and total matter density fields are the best predictors of the WDM particle mass, various gas properties including the electron number density, neutral hydrogen density, temperature, and pressure also contain predictive power. See Section 4.2.

Ultimately, the DREAMS project will consist of thousands of cosmological hydrodynamic simulations that cover a range of systems from galaxy clusters to ultra-faint dwarfs. The long-term goals of the project are to better constrain the properties of DM, understand the dependence of galaxy formation on variations in the dark sector, and understand whether changes to baryonic feedback or dark sector physics can alleviate the small-scale tensions. There are three key extensions that are anticipated for future suites of the DREAMS project. First, is the consideration of CDM and other alternative DM models. Second, is the generalization of the galaxy formation model, both by including more free parameters within TNG and by also extending to other models beyond TNG. Third, is the inclusion of zoom-in simulation suites of different astrophysical targets. Taken together, these simulation suites



will facilitate the comparison of DM predictions with current and upcoming data, enabling robust tests of CDM using small-scale galactic observables.

#### ACKNOWLEDGEMENTS

The DREAMS team thanks the Simons Foundation for their support in hosting and organizing a foundational workshop on the early development of the DREAMS project. JCR acknowledges support from the University of Florida Graduate School’s Graduate Research Fellowship and the Simons Foundation predoctoral program. PT acknowledges support from NSF grant AST-2346977. The work of FVN is supported by the Simons Foundation and by NSF grant AST 2108078. ML, SR, and KEK are supported by the Department of Energy (DOE) under Award Number DE-SC0007968. ML also acknowledges support from the Simons Investigator in Physics Award. MM acknowledges partial support by the National Science Foundation under Grant No. PHY-2010109. FYCR acknowledges the support of program HST-AR-17061.001-A whose support was provided by the National Aeronautical and Space Administration (NASA) through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. DAA acknowledges support by NSF grant AST-2108944, Simons Foundation Award CCA-1018464, and Cottrell Scholar Award CS-CSA-2023-028 by the Research Corporation for Science Advancement. KEK is supported by the NSF Graduate Research Fellowship Program under Grant No. DGE-2039656. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Preliminary work that led to the creation of the DREAMS project utilized Anvil at Purdue University through allocation PHY220051 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

#### DATA AVAILABILITY

The data and code used to produce this paper can be made available upon reasonable request to the corresponding author.

## REFERENCES

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. 2019, in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
- Alam, S., Ata, M., Bailey, S., et al. 2017, *MNRAS*, 470, 2617, doi: [10.1093/mnras/stx721](https://doi.org/10.1093/mnras/stx721)
- Alexander, S., Gleyzer, S., McDonough, E., Toomey, M. W., & Usai, E. 2020, *ApJ*, 893, 15, doi: [10.3847/1538-4357/ab7925](https://doi.org/10.3847/1538-4357/ab7925)
- Anbajagane, D., Evrard, A. E., & Farahi, A. 2022, *MNRAS*, 509, 3441, doi: [10.1093/mnras/stab3177](https://doi.org/10.1093/mnras/stab3177)
- Arun, K., Gudennavar, S. B., & Sivaram, C. 2017, *Advances in Space Research*, 60, 166, doi: [10.1016/j.asr.2017.03.043](https://doi.org/10.1016/j.asr.2017.03.043)
- Bagla, J. S. 2002, *Journal of Astrophysics and Astronomy*, 23, 185, doi: [10.1007/BF02702282](https://doi.org/10.1007/BF02702282)
- Balberg, S., & Shapiro, S. L. 2002, *PhRvL*, 88, 101301, doi: [10.1103/PhysRevLett.88.101301](https://doi.org/10.1103/PhysRevLett.88.101301)
- Banerjee, A., Das, S., Maharana, A., Nadler, E. O., & Sharma, R. K. 2023, *PhRvD*, 108, 043518, doi: [10.1103/PhysRevD.108.043518](https://doi.org/10.1103/PhysRevD.108.043518)
- Banik, N., Bovy, J., Bertone, G., Erkal, D., & de Boer, T. J. L. 2021, *JCAP*, 2021, 043, doi: [10.1088/1475-7516/2021/10/043](https://doi.org/10.1088/1475-7516/2021/10/043)
- Baron, D. 2019, *arXiv e-prints*, arXiv:1904.07248, doi: [10.48550/arXiv.1904.07248](https://doi.org/10.48550/arXiv.1904.07248)
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, *ApJ*, 770, 57, doi: [10.1088/0004-637X/770/1/57](https://doi.org/10.1088/0004-637X/770/1/57)
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013b, *ApJ*, 762, 109, doi: [10.1088/0004-637X/762/2/109](https://doi.org/10.1088/0004-637X/762/2/109)
- Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013c, *ApJ*, 763, 18, doi: [10.1088/0004-637X/763/1/18](https://doi.org/10.1088/0004-637X/763/1/18)
- Bertone, G., & Hooper, D. 2018, *Reviews of Modern Physics*, 90, 045002, doi: [10.1103/RevModPhys.90.045002](https://doi.org/10.1103/RevModPhys.90.045002)
- Bertone, G., & Tait, T. M. P. 2018, *Nature*, 562, 51, doi: [10.1038/s41586-018-0542-z](https://doi.org/10.1038/s41586-018-0542-z)
- Birkinshaw, M. 1999, *Phys. Rept.*, 310, 97, doi: [10.1016/S0370-1573\(98\)00080-5](https://doi.org/10.1016/S0370-1573(98)00080-5)
- Bland-Hawthorn, J., & Gerhard, O. 2016, *ARA&A*, 54, 529, doi: [10.1146/annurev-astro-081915-023441](https://doi.org/10.1146/annurev-astro-081915-023441)
- Bodas, A., Buen-Abad, M. A., Hook, A., & Sundrum, R. 2024, *arXiv e-prints*, arXiv:2401.12286, doi: [10.48550/arXiv.2401.12286](https://doi.org/10.48550/arXiv.2401.12286)
- Bode, P., Ostriker, J. P., & Turok, N. 2001, *ApJ*, 556, 93, doi: [10.1086/321541](https://doi.org/10.1086/321541)
- Böhm, C., Riazuelo, A., Hansen, S. H., & Schaeffer, R. 2002, *PhRvD*, 66, 083505, doi: [10.1103/PhysRevD.66.083505](https://doi.org/10.1103/PhysRevD.66.083505)
- Bohr, S., Zavala, J., Cyr-Racine, F.-Y., et al. 2020, *MNRAS*, 498, 3403, doi: [10.1093/mnras/staa2579](https://doi.org/10.1093/mnras/staa2579)
- Bondi, H., & Hoyle, F. 1944, *MNRAS*, 104, 273, doi: [10.1093/mnras/104.5.273](https://doi.org/10.1093/mnras/104.5.273)
- Bonnivard, V., Combet, C., Daniel, M., et al. 2015, *MNRAS*, 453, 849, doi: [10.1093/mnras/stv1601](https://doi.org/10.1093/mnras/stv1601)
- Bose, S., Deason, A. J., & Frenk, C. S. 2018, *ApJ*, 863, 123, doi: [10.3847/1538-4357/aacbc4](https://doi.org/10.3847/1538-4357/aacbc4)
- Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, *MNRAS*, 415, L40, doi: [10.1111/j.1745-3933.2011.01074.x](https://doi.org/10.1111/j.1745-3933.2011.01074.x)
- . 2012, *MNRAS*, 422, 1203, doi: [10.1111/j.1365-2966.2012.20695.x](https://doi.org/10.1111/j.1365-2966.2012.20695.x)
- Brandt, T. D. 2016, *ApJL*, 824, L31, doi: [10.3847/2041-8205/824/2/L31](https://doi.org/10.3847/2041-8205/824/2/L31)
- Brooks, A. M., Kuhlen, M., Zolotov, A., & Hooper, D. 2013, *ApJ*, 765, 22, doi: [10.1088/0004-637X/765/1/22](https://doi.org/10.1088/0004-637X/765/1/22)
- Brown, S. T., Fattahi, A., McCarthy, I. G., et al. 2024, *arXiv e-prints*, arXiv:2403.11692, doi: [10.48550/arXiv.2403.11692](https://doi.org/10.48550/arXiv.2403.11692)
- Bryan, G. L., Norman, M. L., O’Shea, B. W., et al. 2014, *ApJS*, 211, 19, doi: [10.1088/0067-0049/211/2/19](https://doi.org/10.1088/0067-0049/211/2/19)
- Carlsten, S. G., Greene, J. E., Beaton, R. L., Danieli, S., & Greco, J. P. 2022, *ApJ*, 933, 47, doi: [10.3847/1538-4357/ac6fd7](https://doi.org/10.3847/1538-4357/ac6fd7)
- Cautun, M., Benítez-Llambay, A., Deason, A. J., et al. 2020, *MNRAS*, 494, 4291, doi: [10.1093/mnras/staa1017](https://doi.org/10.1093/mnras/staa1017)
- Chan, T. K., Kereš, D., Oñorbe, J., et al. 2015, *MNRAS*, 454, 2981, doi: [10.1093/mnras/stv2165](https://doi.org/10.1093/mnras/stv2165)
- Chang, J. H., Egana-Ugrinovic, D., Essig, R., & Kouvaris, C. 2019, *JCAP*, 2019, 036, doi: [10.1088/1475-7516/2019/03/036](https://doi.org/10.1088/1475-7516/2019/03/036)
- Corasaniti, P. S., Agarwal, S., Marsh, D. J. E., & Das, S. 2017, *Phys. Rev. D*, 95, 083512, doi: [10.1103/PhysRevD.95.083512](https://doi.org/10.1103/PhysRevD.95.083512)
- Correa, C., Schaller, M., Schaye, J., et al. 2024, *arXiv e-prints*, arXiv:2403.09186, doi: [10.48550/arXiv.2403.09186](https://doi.org/10.48550/arXiv.2403.09186)
- Cyr-Racine, F.-Y., & Sigurdson, K. 2013, *Phys. Rev. D*, 87, 103515, doi: [10.1103/PhysRevD.87.103515](https://doi.org/10.1103/PhysRevD.87.103515)
- Cyr-Racine, F.-Y., Sigurdson, K., Zavala, J., et al. 2016, *PhRvD*, 93, 123527, doi: [10.1103/PhysRevD.93.123527](https://doi.org/10.1103/PhysRevD.93.123527)
- Dalal, N., & Kravtsov, A. 2022, *Phys. Rev. D*, 106, 063517, doi: [10.1103/PhysRevD.106.063517](https://doi.org/10.1103/PhysRevD.106.063517)
- Davis, S. W., & Laor, A. 2011, *The Astrophysical Journal*, 728, 98, doi: [10.1088/0004-637x/728/2/98](https://doi.org/10.1088/0004-637x/728/2/98)
- Dawoodbhoy, T., Shapiro, P. R., Ocvirk, P., et al. 2018, *MNRAS*, 480, 1740, doi: [10.1093/mnras/sty1945](https://doi.org/10.1093/mnras/sty1945)
- de Santi, N. S. M., Shao, H., Villaescusa-Navarro, F., et al. 2023a, *ApJ*, 952, 69, doi: [10.3847/1538-4357/acd1e2](https://doi.org/10.3847/1538-4357/acd1e2)
- . 2023b, *ApJ*, 952, 69, doi: [10.3847/1538-4357/acd1e2](https://doi.org/10.3847/1538-4357/acd1e2)
- Di Cintio, A., Brook, C. B., Macciò, A. V., et al. 2014, *MNRAS*, 437, 415, doi: [10.1093/mnras/stt1891](https://doi.org/10.1093/mnras/stt1891)
- Diaz Rivero, A., & Dvorkin, C. 2020, *PhRvD*, 101, 023515, doi: [10.1103/PhysRevD.101.023515](https://doi.org/10.1103/PhysRevD.101.023515)
- Dome, T., Fialkov, A., Mocz, P., et al. 2022, *Mon. Not. Roy. Astron. Soc.*, 519, 4183, doi: [10.1093/mnras/stac3766](https://doi.org/10.1093/mnras/stac3766)
- Eadie, G., & Jurić, M. 2019, *ApJ*, 875, 159, doi: [10.3847/1538-4357/ab0f97](https://doi.org/10.3847/1538-4357/ab0f97)
- Engler, C., Pillepich, A., Pasquali, A., et al. 2021, *MNRAS*, 507, 4211, doi: [10.1093/mnras/stab2437](https://doi.org/10.1093/mnras/stab2437)

- Enzi, W., Murgia, R., Newton, O., et al. 2021, MNRAS, 506, 5848, doi: [10.1093/mnras/stab1960](https://doi.org/10.1093/mnras/stab1960)
- Essig, R., McDermott, S. D., Yu, H.-B., & Zhong, Y.-M. 2019, PhRvL, 123, 121102, doi: [10.1103/PhysRevLett.123.121102](https://doi.org/10.1103/PhysRevLett.123.121102)
- Fan, J., Katz, A., Randall, L., & Reece, M. 2013, Physics of the Dark Universe, 2, 139, doi: [10.1016/j.dark.2013.07.001](https://doi.org/10.1016/j.dark.2013.07.001)
- Feldmann, R., Quataert, E., Faucher-Giguère, C.-A., et al. 2023, MNRAS, 522, 3831, doi: [10.1093/mnras/stad1205](https://doi.org/10.1093/mnras/stad1205)
- Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, Astron. Astrophys. Rev., 20, 54, doi: [10.1007/s00159-012-0054-z](https://doi.org/10.1007/s00159-012-0054-z)
- Fluri, J., Kacprzak, T., Lucchi, A., et al. 2019, PhRvD, 100, 063514, doi: [10.1103/PhysRevD.100.063514](https://doi.org/10.1103/PhysRevD.100.063514)
- Font, A. S., McCarthy, I. G., Belokurov, V., Brown, S. T., & Stafford, S. G. 2022, MNRAS, 511, 1544, doi: [10.1093/mnras/stac183](https://doi.org/10.1093/mnras/stac183)
- Garrison, L. H., Eisenstein, D. J., Ferrer, D., Maksimova, N. A., & Pinto, P. A. 2021, MNRAS, 508, 575, doi: [10.1093/mnras/stab2482](https://doi.org/10.1093/mnras/stab2482)
- Garrison-Kimmel, S., Boylan-Kolchin, M., Bullock, J. S., & Kirby, E. N. 2014, MNRAS, 444, 222, doi: [10.1093/mnras/stu1477](https://doi.org/10.1093/mnras/stu1477)
- Garrison-Kimmel, S., Hopkins, P. F., Wetzel, A., et al. 2019, MNRAS, 487, 1380, doi: [10.1093/mnras/stz1317](https://doi.org/10.1093/mnras/stz1317)
- Gebhardt, M., Anglés-Alcázar, D., Borrow, J., et al. 2023, arXiv e-prints, arXiv:2307.11832, doi: [10.48550/arXiv.2307.11832](https://doi.org/10.48550/arXiv.2307.11832)
- Geha, M., Wechsler, R. H., Mao, Y.-Y., et al. 2017, ApJ, 847, 4, doi: [10.3847/1538-4357/aa8626](https://doi.org/10.3847/1538-4357/aa8626)
- Gemmell, C., Roy, S., Shen, X., et al. 2023, arXiv e-prints, arXiv:2311.02148, doi: [10.48550/arXiv.2311.02148](https://doi.org/10.48550/arXiv.2311.02148)
- Ghalsasi, A., & McQuinn, M. 2018, PhRvD, 97, 123018, doi: [10.1103/PhysRevD.97.123018](https://doi.org/10.1103/PhysRevD.97.123018)
- Gilman, D., Birrer, S., Nierenberg, A., et al. 2020, Mon. Not. Roy. Astron. Soc., 491, 6077, doi: [10.1093/mnras/stz3480](https://doi.org/10.1093/mnras/stz3480)
- Governato, F., Brook, C., Mayer, L., et al. 2010, Nature, 463, 203, doi: [10.1038/nature08640](https://doi.org/10.1038/nature08640)
- Grand, R. J. J., Gómez, F. A., Marinacci, F., et al. 2017, MNRAS, 467, 179, doi: [10.1093/mnras/stx071](https://doi.org/10.1093/mnras/stx071)
- Greengard, L., & Rokhlin, V. 1987, Journal of Computational Physics, 73, 325, doi: [10.1016/0021-9991\(87\)90140-9](https://doi.org/10.1016/0021-9991(87)90140-9)
- Gupta, A., Zorrilla Matilla, J. M., Hsu, D., & Haiman, Z. 2018, PhRvD, 97, 103515, doi: [10.1103/PhysRevD.97.103515](https://doi.org/10.1103/PhysRevD.97.103515)
- Guzmán, R., GómezFlechoso, M., Martínez-Delgado, D., et al. 2022, *ARRAKIHS Proposal*
- Haardt, F., & Madau, P. 2012, Astrophys. J., 746, 125, doi: [10.1088/0004-637X/746/2/125](https://doi.org/10.1088/0004-637X/746/2/125)
- Hahn, O., & Abel, T. 2011, MNRAS, 415, 2101, doi: [10.1111/j.1365-2966.2011.18820.x](https://doi.org/10.1111/j.1365-2966.2011.18820.x)
- Harnois-Déraps, J., Pen, U.-L., Iliev, I. T., et al. 2013, MNRAS, 436, 540, doi: [10.1093/mnras/stt1591](https://doi.org/10.1093/mnras/stt1591)
- Hausammann, L., Revaz, Y., & Jablonka, P. 2019, A&A, 624, A11, doi: [10.1051/0004-6361/201834871](https://doi.org/10.1051/0004-6361/201834871)
- Hernquist, L., & Katz, N. 1989, ApJS, 70, 419, doi: [10.1086/191344](https://doi.org/10.1086/191344)
- Ho, M., Soltis, J., Farahi, A., et al. 2023, MNRAS, 524, 3289, doi: [10.1093/mnras/stad2005](https://doi.org/10.1093/mnras/stad2005)
- Hockney, R. W., & Eastwood, J. W. 1988, Computer simulation using particles
- Hogan, C. J., & Dalcanton, J. J. 2000, PhRvD, 62, 063511, doi: [10.1103/PhysRevD.62.063511](https://doi.org/10.1103/PhysRevD.62.063511)
- Hoof, S., Geringer-Sameth, A., & Trotta, R. 2020, JCAP, 2020, 012, doi: [10.1088/1475-7516/2020/02/012](https://doi.org/10.1088/1475-7516/2020/02/012)
- Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800, doi: [10.1093/mnras/sty1690](https://doi.org/10.1093/mnras/sty1690)
- Hsueh, J. W., Enzi, W., Vegetti, S., et al. 2020, MNRAS, 492, 3047, doi: [10.1093/mnras/stz3177](https://doi.org/10.1093/mnras/stz3177)
- Ichimaru, S. 1977, ApJ, 214, 840, doi: [10.1086/155314](https://doi.org/10.1086/155314)
- Iršič, V., Viel, M., Haehnelt, M. G., et al. 2017, PhRvD, 96, 023522, doi: [10.1103/PhysRevD.96.023522](https://doi.org/10.1103/PhysRevD.96.023522)
- Jeffrey, N., & Wandelt, B. D. 2020, arXiv e-prints, arXiv:2011.05991, doi: [10.48550/arXiv.2011.05991](https://doi.org/10.48550/arXiv.2011.05991)
- Kafle, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2014, ApJ, 794, 59, doi: [10.1088/0004-637X/794/1/59](https://doi.org/10.1088/0004-637X/794/1/59)
- Kaplan, D. E., Krnjaic, G. Z., Rehermann, K. R., & Wells, C. M. 2010, JCAP, 05, 021, doi: [10.1088/1475-7516/2010/05/021](https://doi.org/10.1088/1475-7516/2010/05/021)
- Kaplinghat, M., Tulin, S., & Yu, H.-B. 2016, Phys. Rev. Lett., 116, 041302, doi: [10.1103/PhysRevLett.116.041302](https://doi.org/10.1103/PhysRevLett.116.041302)
- Katz, H., Ramsoy, M., Rosdahl, J., et al. 2020, MNRAS, 494, 2200, doi: [10.1093/mnras/staa639](https://doi.org/10.1093/mnras/staa639)
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82, doi: [10.1086/307643](https://doi.org/10.1086/307643)
- Koda, J., & Shapiro, P. R. 2011, MNRAS, 415, 1125, doi: [10.1111/j.1365-2966.2011.18684.x](https://doi.org/10.1111/j.1365-2966.2011.18684.x)
- Kulkarni, M., Bryan, G. L., & Visbal, E. 2022, in American Astronomical Society Meeting Abstracts, Vol. 54, American Astronomical Society Meeting #240, 207.02
- Kuzio de Naray, R., Martinez, G. D., Bullock, J. S., & Kaplinghat, M. 2010, ApJL, 710, L161, doi: [10.1088/2041-8205/710/2/L161](https://doi.org/10.1088/2041-8205/710/2/L161)
- Laguë, A., Schwabe, B., Hložek, R., Marsh, D. J. E., & Rogers, K. K. 2024, Phys. Rev. D, 109, 043507, doi: [10.1103/PhysRevD.109.043507](https://doi.org/10.1103/PhysRevD.109.043507)
- Lancaster, L., Giovanetti, C., Mocz, P., et al. 2020, JCAP, 01, 001, doi: [10.1088/1475-7516/2020/01/001](https://doi.org/10.1088/1475-7516/2020/01/001)
- Lazar, A., Bullock, J. S., Boylan-Kolchin, M., et al. 2020, MNRAS, 497, 2393, doi: [10.1093/mnras/staa2101](https://doi.org/10.1093/mnras/staa2101)
- Lelli, F., McGaugh, S. S., & Schombert, J. M. 2016, AJ, 152, 157, doi: [10.3847/0004-6256/152/6/157](https://doi.org/10.3847/0004-6256/152/6/157)
- Leo, M., Baugh, C. M., Li, B., & Pascoli, S. 2017, JCAP, 2017, 017, doi: [10.1088/1475-7516/2017/11/017](https://doi.org/10.1088/1475-7516/2017/11/017)
- Li, X., Hui, L., & Bryan, G. L. 2019, Phys. Rev. D, 99, 063509, doi: [10.1103/PhysRevD.99.063509](https://doi.org/10.1103/PhysRevD.99.063509)



- Lora, V., Magana, J., Bernal, A., Sanchez-Salcedo, F. J., & Grebel, E. K. 2012, *JCAP*, 02, 011, doi: [10.1088/1475-7516/2012/02/011](https://doi.org/10.1088/1475-7516/2012/02/011)
- Lovell, M. R., Frenk, C. S., Eke, V. R., et al. 2014, *MNRAS*, 439, 300, doi: [10.1093/mnras/stt2431](https://doi.org/10.1093/mnras/stt2431)
- Lovell, M. R., Zavala, J., Vogelsberger, M., et al. 2018, *MNRAS*, 477, 2886, doi: [10.1093/mnras/sty818](https://doi.org/10.1093/mnras/sty818)
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201, doi: [10.48550/arXiv.0912.0201](https://doi.org/10.48550/arXiv.0912.0201)
- Luo, Y., Leauthaud, A., Greene, J., et al. 2023, arXiv e-prints, arXiv:2305.19310, doi: [10.48550/arXiv.2305.19310](https://doi.org/10.48550/arXiv.2305.19310)
- Macciò, A. V., Paduoiu, S., Anderhalden, D., Schneider, A., & Moore, B. 2012, *MNRAS*, 424, 1105, doi: [10.1111/j.1365-2966.2012.21284.x](https://doi.org/10.1111/j.1365-2966.2012.21284.x)
- Mace, C., Carton Zeng, Z., Peter, A. H. G., et al. 2024, arXiv e-prints, arXiv:2402.01604, doi: [10.48550/arXiv.2402.01604](https://doi.org/10.48550/arXiv.2402.01604)
- Marinacci, F., Sales, L. V., Vogelsberger, M., Torrey, P., & Springel, V. 2019, *MNRAS*, 489, 4233, doi: [10.1093/mnras/stz2391](https://doi.org/10.1093/mnras/stz2391)
- Mocz, P., Vogelsberger, M., Robles, V. H., et al. 2017, *Mon. Not. Roy. Astron. Soc.*, 471, 4559, doi: [10.1093/mnras/stx1887](https://doi.org/10.1093/mnras/stx1887)
- Mocz, P., et al. 2019, *Phys. Rev. Lett.*, 123, 141301, doi: [10.1103/PhysRevLett.123.141301](https://doi.org/10.1103/PhysRevLett.123.141301)
- Monaghan, J. J. 1992, *ARA&A*, 30, 543, doi: [10.1146/annurev.aa.30.090192.002551](https://doi.org/10.1146/annurev.aa.30.090192.002551)
- Moore, B., Ghigna, S., Governato, F., et al. 1999, *ApJL*, 524, L19, doi: [10.1086/312287](https://doi.org/10.1086/312287)
- Mroczkowski, T., et al. 2019, *Space Sci. Rev.*, 215, 17, doi: [10.1007/s11214-019-0581-2](https://doi.org/10.1007/s11214-019-0581-2)
- Nadler, E. O., Birrer, S., Gilman, D., et al. 2021a, *ApJ*, 917, 7, doi: [10.3847/1538-4357/abf9a3](https://doi.org/10.3847/1538-4357/abf9a3)
- Nadler, E. O., Wechsler, R. H., Bechtol, K., et al. 2020, *ApJ*, 893, 48, doi: [10.3847/1538-4357/ab846a](https://doi.org/10.3847/1538-4357/ab846a)
- Nadler, E. O., Drlica-Wagner, A., Bechtol, K., et al. 2021b, *PhRvL*, 126, 091101, doi: [10.1103/PhysRevLett.126.091101](https://doi.org/10.1103/PhysRevLett.126.091101)
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 477, 1206, doi: [10.1093/mnras/sty618](https://doi.org/10.1093/mnras/sty618)
- Nelson, D., Pillepich, A., Springel, V., et al. 2019, *MNRAS*, 490, 3234, doi: [10.1093/mnras/stz2306](https://doi.org/10.1093/mnras/stz2306)
- Ni, Y., Genel, S., Anglés-Alcázar, D., et al. 2023, *ApJ*, 959, 136, doi: [10.3847/1538-4357/ad022a](https://doi.org/10.3847/1538-4357/ad022a)
- Nishikawa, H., Boddy, K. K., & Kaplinghat, M. 2020, *PhRvD*, 101, 063009, doi: [10.1103/PhysRevD.101.063009](https://doi.org/10.1103/PhysRevD.101.063009)
- Ntampaka, M., ZuHone, J., Eisenstein, D., et al. 2019, *ApJ*, 876, 82, doi: [10.3847/1538-4357/ab14eb](https://doi.org/10.3847/1538-4357/ab14eb)
- Oñorbe, J., Garrison-Kimmel, S., Maller, A. H., et al. 2014, *MNRAS*, 437, 1894, doi: [10.1093/mnras/stt2020](https://doi.org/10.1093/mnras/stt2020)
- Obreja, A., Stinson, G. S., Dutton, A. A., et al. 2016, *MNRAS*, 459, 467, doi: [10.1093/mnras/stw690](https://doi.org/10.1093/mnras/stw690)
- Oman, K. A., Navarro, J. F., Fattahi, A., et al. 2015, *MNRAS*, 452, 3650, doi: [10.1093/mnras/stv1504](https://doi.org/10.1093/mnras/stv1504)
- O’Neil, S., Vogelsberger, M., Heeba, S., et al. 2023, *MNRAS*, 524, 288, doi: [10.1093/mnras/stad1850](https://doi.org/10.1093/mnras/stad1850)
- Ouchi, M., Ono, Y., & Shibuya, T. 2020, *Ann. Rev. Astron. Astrophys.*, 58, 617, doi: [10.1146/annurev-astro-032620-021859](https://doi.org/10.1146/annurev-astro-032620-021859)
- Painter, C. A., Boylan-Kolchin, M., Mocz, P., & Vogelsberger, M. 2024, arXiv e-prints, arXiv:2402.16945, doi: [10.48550/arXiv.2402.16945](https://doi.org/10.48550/arXiv.2402.16945)
- Palubski, I., Slone, O., Kaplinghat, M., Lisanti, M., & Jiang, F. 2024, arXiv e-prints, arXiv:2402.12452, doi: [10.48550/arXiv.2402.12452](https://doi.org/10.48550/arXiv.2402.12452)
- Papastergis, E., Giovanelli, R., Haynes, M. P., & Shankar, F. 2015, *A&A*, 574, A113, doi: [10.1051/0004-6361/201424909](https://doi.org/10.1051/0004-6361/201424909)
- Pawlowski, M. S. 2018, *Modern Physics Letters A*, 33, 1830004, doi: [10.1142/S0217732318300045](https://doi.org/10.1142/S0217732318300045)
- Pawlowski, M. S., Famaey, B., Jerjen, H., et al. 2014, *MNRAS*, 442, 2362, doi: [10.1093/mnras/stu1005](https://doi.org/10.1093/mnras/stu1005)
- Pearson, S., Price-Whelan, A. M., Hogg, D. W., et al. 2022, *ApJ*, 941, 19, doi: [10.3847/1538-4357/ac9bfb](https://doi.org/10.3847/1538-4357/ac9bfb)
- Peel, A., Lalande, F., Starck, J.-L., et al. 2019, *PhRvD*, 100, 023508, doi: [10.1103/PhysRevD.100.023508](https://doi.org/10.1103/PhysRevD.100.023508)
- Peirani, S., Dubois, Y., Volonteri, M., et al. 2017, *MNRAS*, 472, 2153, doi: [10.1093/mnras/stx2099](https://doi.org/10.1093/mnras/stx2099)
- Peirani, S., Sonnenfeld, A., Gavazzi, R., et al. 2019, *MNRAS*, 483, 4615, doi: [10.1093/mnras/sty3475](https://doi.org/10.1093/mnras/sty3475)
- Pillepich, A., Springel, V., Nelson, D., et al. 2018a, *MNRAS*, 473, 4077, doi: [10.1093/mnras/stx2656](https://doi.org/10.1093/mnras/stx2656)
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018b, *MNRAS*, 475, 648, doi: [10.1093/mnras/stx3112](https://doi.org/10.1093/mnras/stx3112)
- Pillepich, A., Nelson, D., Springel, V., et al. 2019, *MNRAS*, 490, 3196, doi: [10.1093/mnras/stz2338](https://doi.org/10.1093/mnras/stz2338)
- Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2016, *A&A*, 594, A11, doi: [10.1051/0004-6361/201526926](https://doi.org/10.1051/0004-6361/201526926)
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Potter, D., Stadel, J., & Teyssier, R. 2017, *Computational Astrophysics and Cosmology*, 4, 2, doi: [10.1186/s40668-017-0021-1](https://doi.org/10.1186/s40668-017-0021-1)
- Pritchard, J. R., & Loeb, A. 2012, *Reports on Progress in Physics*, 75, 086901, doi: [10.1088/0034-4885/75/8/086901](https://doi.org/10.1088/0034-4885/75/8/086901)
- Read, J. I., Iorio, G., Agertz, O., & Fraternali, F. 2017, *MNRAS*, 467, 2019, doi: [10.1093/mnras/stx147](https://doi.org/10.1093/mnras/stx147)
- Reid, B. A., Percival, W. J., Eisenstein, D. J., et al. 2010, *MNRAS*, 404, 60, doi: [10.1111/j.1365-2966.2010.16276.x](https://doi.org/10.1111/j.1365-2966.2010.16276.x)
- Ribli, D., Pataki, B. Á., Zorrilla Matilla, J. M., et al. 2019, *MNRAS*, 490, 1843, doi: [10.1093/mnras/stz2610](https://doi.org/10.1093/mnras/stz2610)
- Robertson, A., Harvey, D., Massey, R., et al. 2019, *MNRAS*, 488, 3646, doi: [10.1093/mnras/stz1815](https://doi.org/10.1093/mnras/stz1815)
- Robles, V. H., Bullock, J. S., Elbert, O. D., et al. 2017, *MNRAS*, 472, 2945, doi: [10.1093/mnras/stx2253](https://doi.org/10.1093/mnras/stx2253)

- Roca-Fàbrega, S., Kim, J.-H., Hausammann, L., et al. 2021, *ApJ*, 917, 64, doi: [10.3847/1538-4357/ac088a](https://doi.org/10.3847/1538-4357/ac088a)
- Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, *MNRAS*, 449, 49, doi: [10.1093/mnras/stv264](https://doi.org/10.1093/mnras/stv264)
- Rose, J. C., Torrey, P., Vogelsberger, M., & O’Neil, S. 2023, *MNRAS*, 519, 5623, doi: [10.1093/mnras/stac3634](https://doi.org/10.1093/mnras/stac3634)
- Rose, J. C., Torrey, P., Villaescusa-Navarro, F., et al. 2024, *MNRAS*, 527, 739, doi: [10.1093/mnras/stad3260](https://doi.org/10.1093/mnras/stad3260)
- Roy, S., Shen, X., Lisanti, M., et al. 2023, *Astrophys. J. Lett.*, 954, L40, doi: [10.3847/2041-8213/ace2c8](https://doi.org/10.3847/2041-8213/ace2c8)
- Ryan, M., Shandera, S., Gurian, J., & Jeong, D. 2022, *ApJ*, 934, 122, doi: [10.3847/1538-4357/ac75e5](https://doi.org/10.3847/1538-4357/ac75e5)
- Ryu, D., Kang, H., Cho, J., & Das, S. 2008, *Science*, 320, 909, doi: [10.1126/science.1154923](https://doi.org/10.1126/science.1154923)
- Sabti, N., Muñoz, J. B., & Blas, D. 2022, *Astrophys. J. Lett.*, 928, L20, doi: [10.3847/2041-8213/ac5e9c](https://doi.org/10.3847/2041-8213/ac5e9c)
- Sales, L. V., Wetzel, A., & Fattahi, A. 2022, *Nature Astronomy*, 6, 897, doi: [10.1038/s41550-022-01689-w](https://doi.org/10.1038/s41550-022-01689-w)
- Sameie, O., Boylan-Kolchin, M., Sanderson, R., et al. 2021, *MNRAS*, 507, 720, doi: [10.1093/mnras/stab2173](https://doi.org/10.1093/mnras/stab2173)
- Samuel, J., Wetzel, A., Santistevan, I., et al. 2022, *MNRAS*, 514, 5276, doi: [10.1093/mnras/stac1706](https://doi.org/10.1093/mnras/stac1706)
- Sawala, T., Cautun, M., Frenk, C., et al. 2023, *Nature Astronomy*, 7, 481, doi: [10.1038/s41550-022-01856-z](https://doi.org/10.1038/s41550-022-01856-z)
- Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, *Mon. Not. Roy. Astron. Soc.*, 318, 817, doi: [10.1046/j.1365-8711.2000.03815.x](https://doi.org/10.1046/j.1365-8711.2000.03815.x)
- Schuldt, S., Suyu, S. H., Meinhardt, T., et al. 2021, *A&A*, 646, A126, doi: [10.1051/0004-6361/202039574](https://doi.org/10.1051/0004-6361/202039574)
- Schultz, C., Oñorbe, J., Abazajian, K. N., & Bullock, J. S. 2014, *MNRAS*, 442, 1597, doi: [10.1093/mnras/stu976](https://doi.org/10.1093/mnras/stu976)
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337
- Shao, H., Villaescusa-Navarro, F., Villanueva-Domingo, P., et al. 2022, arXiv e-prints, arXiv:2209.06843, doi: [10.48550/arXiv.2209.06843](https://doi.org/10.48550/arXiv.2209.06843)
- Shao, H., de Santi, N. S. M., Villaescusa-Navarro, F., et al. 2023, *ApJ*, 956, 149, doi: [10.3847/1538-4357/acee6f](https://doi.org/10.3847/1538-4357/acee6f)
- Shapiro, S. L., Lightman, A. P., & Eardley, D. M. 1976, *ApJ*, 204, 187, doi: [10.1086/154162](https://doi.org/10.1086/154162)
- Shen, X., Hopkins, P., Necib, L., et al. 2022a, in *American Astronomical Society Meeting Abstracts*, Vol. 54, American Astronomical Society Meeting #240, 347.06
- Shen, X., Hopkins, P. F., Necib, L., et al. 2021, *MNRAS*, 506, 4421, doi: [10.1093/mnras/stab2042](https://doi.org/10.1093/mnras/stab2042)
- . 2022b, arXiv e-prints, arXiv:2206.05327, doi: [10.48550/arXiv.2206.05327](https://doi.org/10.48550/arXiv.2206.05327)
- Shen, X., Borrow, J., Vogelsberger, M., et al. 2024, *MNRAS*, 527, 2835, doi: [10.1093/mnras/stad3397](https://doi.org/10.1093/mnras/stad3397)
- Silverman, M., Bullock, J. S., Kaplinghat, M., Robles, V. H., & Valli, M. 2023, *MNRAS*, 518, 2418, doi: [10.1093/mnras/stac3232](https://doi.org/10.1093/mnras/stac3232)
- Simon, J. D. 2019, *ARA&A*, 57, 375, doi: [10.1146/annurev-astro-091918-104453](https://doi.org/10.1146/annurev-astro-091918-104453)
- Slone, O., Jiang, F., Lisanti, M., & Kaplinghat, M. 2021, arXiv e-prints, arXiv:2108.03243, doi: [10.48550/arXiv.2108.03243](https://doi.org/10.48550/arXiv.2108.03243)
- Sobol’, I. M. 1967, *Zhurnal Vychislitel’noi Matematiki i Matematicheskoi Fiziki*, 7, 784
- Sohn, S. T., Watkins, L. L., Fardal, M. A., et al. 2018, *ApJ*, 862, 52, doi: [10.3847/1538-4357/aacd0b](https://doi.org/10.3847/1538-4357/aacd0b)
- Sommer-Larsen, J., & Dolgov, A. 2001, *ApJ*, 551, 608, doi: [10.1086/320211](https://doi.org/10.1086/320211)
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757, doi: [10.48550/arXiv.1503.03757](https://doi.org/10.48550/arXiv.1503.03757)
- Spergel, D. N., & Steinhardt, P. J. 2000, *PhRvL*, 84, 3760, doi: [10.1103/PhysRevLett.84.3760](https://doi.org/10.1103/PhysRevLett.84.3760)
- Springel, V. 2005, *MNRAS*, 364, 1105, doi: [10.1111/j.1365-2966.2005.09655.x](https://doi.org/10.1111/j.1365-2966.2005.09655.x)
- . 2010, *MNRAS*, 401, 791, doi: [10.1111/j.1365-2966.2009.15715.x](https://doi.org/10.1111/j.1365-2966.2009.15715.x)
- Springel, V., Pakmor, R., & Weinberger, R. 2019, AREPO: Cosmological magnetohydrodynamical moving-mesh simulation code, *Astrophysics Source Code Library*, record ascl:1909.010, <http://ascl.net/1909.010>
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, 328, 726, doi: [10.1046/j.1365-8711.2001.04912.x](https://doi.org/10.1046/j.1365-8711.2001.04912.x)
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629, doi: [10.1038/nature03597](https://doi.org/10.1038/nature03597)
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, 475, 676, doi: [10.1093/mnras/stx3304](https://doi.org/10.1093/mnras/stx3304)
- Teyssier, R. 2002, *A&A*, 385, 337, doi: [10.1051/0004-6361:20011817](https://doi.org/10.1051/0004-6361:20011817)
- Teyssier, R., Moore, B., Martizzi, D., Dubois, Y., & Mayer, L. 2011, *MNRAS*, 414, 195, doi: [10.1111/j.1365-2966.2011.18399.x](https://doi.org/10.1111/j.1365-2966.2011.18399.x)
- The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro, doi: [10.48550/arXiv.astro-ph/0510346](https://doi.org/10.48550/arXiv.astro-ph/0510346)
- Todoroki, K., & Medvedev, M. V. 2019, *Mon. Not. Roy. Astron. Soc.*, 483, 3983, doi: [10.1093/mnras/sty3401](https://doi.org/10.1093/mnras/sty3401)
- Todoroki, K., & Medvedev, M. V. 2019, *MNRAS*, 483, 4004, doi: [10.1093/mnras/sty3353](https://doi.org/10.1093/mnras/sty3353)
- . 2022, *MNRAS*, 510, 4249, doi: [10.1093/mnras/stab3764](https://doi.org/10.1093/mnras/stab3764)
- Tollerud, E. J., Boylan-Kolchin, M., & Bullock, J. S. 2014, *MNRAS*, 440, 3511, doi: [10.1093/mnras/stu474](https://doi.org/10.1093/mnras/stu474)
- Tollet, E., Macciò, A. V., Dutton, A. A., et al. 2016, *MNRAS*, 456, 3542, doi: [10.1093/mnras/stv2856](https://doi.org/10.1093/mnras/stv2856)
- Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, *MNRAS*, 438, 1985, doi: [10.1093/mnras/stt2295](https://doi.org/10.1093/mnras/stt2295)

- Tulin, S., & Yu, H.-B. 2018, *PhR*, 730, 1,  
doi: [10.1016/j.physrep.2017.11.004](https://doi.org/10.1016/j.physrep.2017.11.004)
- van den Aarssen, L. G., Bringmann, T., & Pfrommer, C. 2012,  
*Phys. Rev. Lett.*, 109, 231301,  
doi: [10.1103/PhysRevLett.109.231301](https://doi.org/10.1103/PhysRevLett.109.231301)
- Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013,  
*PhRvD*, 88, 043502, doi: [10.1103/PhysRevD.88.043502](https://doi.org/10.1103/PhysRevD.88.043502)
- Villaescusa-Navarro, F., Wandelt, B. D., Anglés-Alcázar, D., et al.  
2022a, *ApJ*, 928, 44, doi: [10.3847/1538-4357/ac54a5](https://doi.org/10.3847/1538-4357/ac54a5)
- Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al.  
2021a, *ApJ*, 915, 71, doi: [10.3847/1538-4357/abf7ba](https://doi.org/10.3847/1538-4357/abf7ba)
- Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., et al.  
2021b, *arXiv e-prints*, *arXiv*:2109.10360,  
doi: [10.48550/arXiv.2109.10360](https://doi.org/10.48550/arXiv.2109.10360)
- Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al.  
2021c, *arXiv e-prints*, *arXiv*:2109.09747,  
doi: [10.48550/arXiv.2109.09747](https://doi.org/10.48550/arXiv.2109.09747)
- Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., et al.  
2022b, *ApJS*, 259, 61, doi: [10.3847/1538-4365/ac5ab0](https://doi.org/10.3847/1538-4365/ac5ab0)
- Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2022, *ApJ*,  
937, 115, doi: [10.3847/1538-4357/ac8930](https://doi.org/10.3847/1538-4357/ac8930)
- Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, *MNRAS*, 436,  
3031, doi: [10.1093/mnras/stt1789](https://doi.org/10.1093/mnras/stt1789)
- . 2014a, *MNRAS*, 438, 3607, doi: [10.1093/mnras/stt2414](https://doi.org/10.1093/mnras/stt2414)
- Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020,  
*Nature Reviews Physics*, 2, 42, doi: [10.1038/s42254-019-0127-2](https://doi.org/10.1038/s42254-019-0127-2)
- Vogelsberger, M., Zavala, J., Cyr-Racine, F.-Y., et al. 2016,  
*MNRAS*, 460, 1399, doi: [10.1093/mnras/stw1076](https://doi.org/10.1093/mnras/stw1076)
- Vogelsberger, M., Zavala, J., Schutz, K., & Slatyer, T. R. 2019,  
*MNRAS*, 484, 5437, doi: [10.1093/mnras/stz340](https://doi.org/10.1093/mnras/stz340)
- Vogelsberger, M., Zavala, J., Simpson, C., & Jenkins, A. 2014b,  
*MNRAS*, 444, 3684, doi: [10.1093/mnras/stu1713](https://doi.org/10.1093/mnras/stu1713)
- Wang, J., & White, S. D. M. 2007, *MNRAS*, 380, 93,  
doi: [10.1111/j.1365-2966.2007.12053.x](https://doi.org/10.1111/j.1365-2966.2007.12053.x)
- Watkins, L. L., Evans, N. W., & An, J. H. 2010, *MNRAS*, 406, 264,  
doi: [10.1111/j.1365-2966.2010.16708.x](https://doi.org/10.1111/j.1365-2966.2010.16708.x)
- Weinberger, R., Springel, V., & Pakmor, R. 2020, *ApJS*, 248, 32,  
doi: [10.3847/1538-4365/ab908c](https://doi.org/10.3847/1538-4365/ab908c)
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, *MNRAS*,  
465, 3291, doi: [10.1093/mnras/stw2944](https://doi.org/10.1093/mnras/stw2944)
- Weinberger, R., Springel, V., Pakmor, R., et al. 2018, *MNRAS*,  
479, 4056, doi: [10.1093/mnras/sty1733](https://doi.org/10.1093/mnras/sty1733)
- Wilkinson, M. I., & Evans, N. W. 1999, *MNRAS*, 310, 645,  
doi: [10.1046/j.1365-8711.1999.02964.x](https://doi.org/10.1046/j.1365-8711.1999.02964.x)
- Zeng, Z. C., Peter, A. H. G., Du, X., et al. 2022, *MNRAS*, 513,  
4845, doi: [10.1093/mnras/stac1094](https://doi.org/10.1093/mnras/stac1094)
- Zhang, X., Wang, Y., Zhang, W., et al. 2019, *arXiv e-prints*,  
*arXiv*:1902.05965, doi: [10.48550/arXiv.1902.05965](https://doi.org/10.48550/arXiv.1902.05965)



## APPENDIX

## A. ZOOM-IN PROCEDURE

This work presents a suite of Milky Way (MW) zoom-in simulations. Creating this set of simulations is a multi-step process. In brief, the procedure is to: (1) simulate a uniform box with no baryons at a low resolution, (2) perform a zoom-in simulation with no baryons at an intermediate resolution, and (3) then perform a final zoom-in with baryons at the fiducial resolution. This Appendix provides greater detail for each of these steps.

For the first step, the uniform-box simulation has a  $(100 h^{-1} \text{ Mpc})^3$  volume with  $256^3$  particles. This provides a mass resolution of  $5 \times 10^9 h^{-1} M_\odot$  and a spatial resolution of  $4.9 h^{-1} \text{ kpc}$  at  $z = 0$ . At this resolution, a MW-mass halo consists of at least 300 particles, which is sufficient to locate the halo in the box, although not good enough to resolve its internal structure. This simulation takes 200 cpu-hours to complete.

For the second step, we choose a random halo from this uniform box whose virial mass falls within  $(1.58\text{--}1.61) \times 10^{12} M_\odot$ . This range is toward the upper end of the observed MW mass estimates compiled in [Bland-Hawthorn & Gerhard \(2016\)](#).<sup>8</sup> We also apply an isolation cut, requiring that there are no other halos more massive than  $7.2 \times 10^{11} h^{-1} M_\odot$  within  $1 h^{-1} \text{ Mpc}$  of the selected halo. This criterion is implemented for computational reasons as halos that are close to more massive halos take too long to simulate or have significant contamination from low-resolution particles within the halo at  $z = 0$ . Overall, this isolation criterion excludes  $\sim 12\%$  of all MW-mass galaxies found in the uniform box. Future work will explore the possibility of removing this isolation criterion so that MW-analogs with large neighbors (like Andromeda) can be studied.

For the target halo, we select all particles within  $5\times$  its virial radius, trace their positions back to  $z = 127$ , and define an ellipse that encompasses them. This ellipse is then used to determine the high-resolution region of the intermediate N-body zoom-in simulation. In this region, the mass resolution is  $1.0 \times 10^7 h^{-1} M_\odot$  and the spatial resolution is  $0.61 h^{-1} \text{ kpc}$  at  $z = 0$ . The target halo contains  $\sim 1 \times 10^5$  particles, which allows for a better definition of its internal structure. The low-resolution region fills the rest of the box. Far from the target halo ( $\gtrsim 10 \text{ Mpc}$  at  $z = 127$ ), the simulation has a mass resolution of  $3 \times 10^{10} h^{-1} M_\odot$ . Surrounding the buffer region, the simulation has concentric regions of decreasing mass that step from the low-resolution region to the high-resolution region. On average, each N-body zoom-in simulation takes 250 cpu-hours, but can take over 1000 cpu-hours if the MW-mass halo is in a crowded environment. As it turns out, this process is more efficient than simulating a uniform box at sufficient resolution to properly identify the internal structure of a target MW-mass halo. As the zoom-in procedure requires over 1000 of these simulations, the increased computation time would be comparable to that spent on the final hydrodynamic suite.

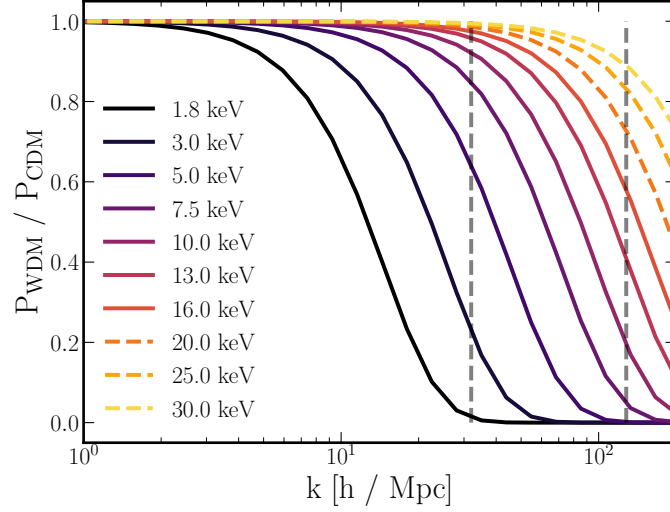
We use the results from the N-body zoom-in simulation from the second step to create the initial conditions for the final hydrodynamic zoom-in simulation. In this process, there can be significant contamination from low-resolution particles in the intermediate N-body zoom-in due to the low resolution of the initial uniform box. To produce a less-contaminated high-resolution region for the final zoom-in simulation, we ensure that we include all particles, high and low resolution, within  $5 \times M_{\text{vir}}$  during this step. We then calculate the ellipse that encloses all particles at  $z = 127$  and use this new ellipse to define the high-resolution region.

For  $\sim 15\%$  of the simulations, this procedure produces initial conditions that can take an order of magnitude longer to simulate than average, due to a large high-resolution region. For these simulations, which we define as having more than  $10^7$  high-resolution particles, we use a convex hull to define the region that encompasses the high-resolution particles. The convex hull allows for a smaller high-resolution region while still encompassing all particles that will be in the MW-mass halo at  $z = 0$ .

[Oñorbe et al. \(2014\)](#) has shown that initial conditions created in convex hulls can result in greater contamination from low-resolution particles than ellipses. Hence, we only use convex hulls when necessary for computational purposes. We do find that some simulations that use the convex hull either result in significant contamination from low-resolution particles ( $>1\%$  by mass within the virial radius) or still maintain a significant computational cost. For these simulations, we implement a new method to reduce the high-resolution Lagrangian region and better target the area around the MW-mass halo.

The method begins by first tracing all particles within the halo's virial radius back to  $z = 127$ . Next, we calculate the distance from each of those particles to the center-of-mass of the high-resolution Lagrangian region at  $z = 127$ . For each particle, we also calculate the distance to its 1024<sup>th</sup> neighbor at  $z = 127$ , only considering the particles that are within the virial radius at  $z = 0$ . We then remove any particle whose distance to the center of the Lagrangian region added to the distance to its 1024<sup>th</sup> neighbor

<sup>8</sup> Due to the low resolution of the uniform box, the mass criterion is not maintained at higher resolution. The average mass of the final MW targets is  $(1.3 \pm 0.2) \times 10^{12} M_\odot$ , still within the acceptable observational range ([Wilkinson & Evans 1999](#); [Watkins et al. 2010](#); [Kafle et al. 2014](#); [Sohn et al. 2018](#); [Eadie & Jurić 2019](#)). In the future, additional care should be taken to choose halos that fall within the desired mass range by checking the isolation criterion after the intermediate N-body zoom-in simulation.



**Figure B1.** The suppression of power relative to CDM for the WDM particle masses used in the DREAMS suites, plotted as a function of comoving wavenumber. The vertical grey-dashed line indicates the Nyquist frequency for the uniform (zoom) simulations at  $32$  ( $128$ )  $h \text{ Mpc}^{-1}$ . The colored dashed lines ( $> 16$  keV) are only included in the zoom-in suite as they are indistinguishable from CDM in the uniform-box suite.

is greater than  $5\times$  the average distance to the center of the Lagrangian region. This cut removes any low-resolution particles that end up within the virial radius of the target halo, but that started from a much further distance. The procedure does not remove most low-resolution particles that are pulled from nearby regions.

After removing the outlier particles, we rebuild the initial conditions around the MW-mass galaxy to maintain a reasonable buffer between the high-resolution and low-resolution regions. To do this, we first trace all particles within  $5\times$  of the target halo's virial radius (the same particles used previously to define the target halo) back to  $z = 127$ . For every particle that passes the cut described above, we find the  $1024^{\text{th}}$  closest particles from those within  $5\times$  the virial radius. Including both the particles that pass the cut and their selected neighbors, we redraw an enclosing ellipse around these particles to make the ICs for the hydrodynamic simulation.

## B. WARM DARK MATTER IMPLEMENTATION

Figure B1 illustrates the suppression of the Warm Dark Matter (WDM) power spectrum relative to Cold Dark Matter (CDM), as a function of comoving wavenumber. Each colored line corresponds to a different WDM mass, starting from 1.8 keV and extending to 30 keV. The vertical dashed gray lines indicate the Nyquist frequencies for the uniform-box ( $32 h \text{ Mpc}^{-1}$ ) and zoom-in ( $128 h \text{ Mpc}^{-1}$ ) WDM simulations—see Tab. 1 for further details. At the uniform-box resolution, WDM masses greater than 16 keV are indistinguishable from CDM and are thus not included in this suite. The zoom-in suite includes masses up to 30 keV, as indicated by the dashed colored lines in Fig. B1, to ensure that inference performed on this suite includes a CDM-like prediction. Note that WDM models with masses  $\lesssim 2$  keV are strongly ruled out by multiple analyses (Iršič et al. 2017; Hsueh et al. 2020; Enzi et al. 2021; Nadler et al. 2021a,b), but are included in the suites to help machine learning models learn the characteristic features that distinguish WDM from CDM.

To simulate thermal-relic WDM, we apply a suppression to the initial matter power spectrum compared to CDM. The relationship between the WDM power spectrum,  $P_{\text{WDM}}(k)$ , and CDM power spectrum,  $P_{\text{CDM}}(k)$ , depends on the wavenumber,  $k$ . Following Bode et al. (2001), we take  $P_{\text{WDM}}(k) = \beta(k)P_{\text{CDM}}(k)$  with

$$\beta(k) = \left( \left( 1 + (\alpha k)^{2.4} \right)^{-5.0/1.2} \right)^2 \quad \text{and} \quad (B1)$$

$$\alpha = 0.048 h^{-1} \text{ Mpc} \left( \frac{M_{\text{WDM}}}{\text{keV}} \right)^{-1.15} \left( \frac{\Omega_m - \Omega_b}{0.4} \right)^{0.15} \left( \frac{h}{0.65} \right)^{1.3}. \quad (B2)$$

Here,  $M_{\text{WDM}}$  is the WDM particle mass in keV,  $\Omega_m$  is the total matter density, and  $\Omega_b$  is the baryon density. The value of  $\Omega_m$  is fixed to 0.302 for the zoom-in suite and is varied according to a Sobol sequence within the range  $\Omega_m \in [0.1, 0.5]$  for the uniform-box suite.

Some simulations of WDM include additional velocity kicks to the simulation particles at early times (e.g., Wang & White 2007; Macciò et al. 2012; Viel et al. 2013). These kicks are meant to approximate the additional thermal velocities of the DM particles. Leo et al. (2017) investigated whether kicks accurately capture the underlying population they are trying to model by examining how including kicks and varying simulation resolution affect power at small scales. They found that including velocity kicks in the simulations introduces greater numerical noise and artificial fragmentation than if they were excluded. Therefore, we do not include thermal kicks in our simulations to reduce numerical noise.

Artificial fragmentation can occur in WDM simulations due to artificial power at small scales present from discrete sampling. During the evolution of the simulation, this artificial power can grow to the size of a small halo—a so-called spurious halo. Using the equation for the halo masses expected from artificial fragmentation from Wang & White (2007), we estimate a maximum mass of  $\sim 8 \times 10^8 h^{-1} M_\odot$  for the spurious halos in the WDM uniform-box simulation (assuming an extreme model with  $\Omega_m = 0.5$  and  $M_{\text{WDM}} = 1.8 \text{ keV}$ ). The mass resolution for the WDM uniform box with  $\Omega_m = 0.5$  is  $1.3 \times 10^8 h^{-1} M_\odot$ . Therefore, the scale of spurious structure formation in this case is  $\sim 6$  particles. Because the image-making procedure smooths the density field over a kernel of 32 particles, the artificial fragmentation should not significantly affect the results.

For the zoom-in simulations, the maximum mass for the spurious halos with 1.8 keV WDM is  $\sim 7 \times 10^7 h^{-1} M_\odot$ . These simulations have a mass resolution of  $1.2 \times 10^6 h^{-1} M_\odot$ , leading to spurious structures of  $\sim 60$  particles. These structures are larger, in total particle count, than those in the uniform-box simulation and may play a more pronounced role in the galaxy evolution for the most extreme WDM models. The analysis in Section 3 only includes galaxies with  $M_{\text{halo}} > 10^8 h^{-1} M_\odot$  and therefore should not be affected by the presence of artificial halos.

### C. MACHINE LEARNING ARCHITECTURES

This section provides details of the machine learning architectures used for the analyses presented in the main body of the paper.

#### C.1. Emulator

Section 3 presents a first application of the WDM zoom-in suite, using emulation to demonstrate how satellite count is affected by WDM particle mass as well as the SN and AGN parameters varied in this study. The emulator consists of multiple layers of fully-connected nodes. The first layer comprises the inputs to the network: WDM particle mass,  $\epsilon_w$ ,  $\kappa_w$ , and  $\epsilon_{f,\text{high}}$ . Recall that the cosmological parameters are kept fixed for the MW zoom-in suite. The first layer is followed by a set of layers, each consisting of neurons that are fully connected to the subsequent layer. The number of neurons in each layer, as well as the total number of layers, are tuned as hyperparameters. The final layer of the network is the outputs of the neural network, which correspond to the mean and standard deviation of MW-mass satellite counts for the example in Section 3.

The emulator is trained on the MW-mass hosts in the suite, which are split into training, validation, and testing sets, each comprising 90%, 5%, and 5% of the total, respectively. We find that increasing the size of the training and testing sets does not significantly affect the results. Training is performed over 500 epochs using the loss function (Villaescusa-Navarro et al. 2022b; Jeffrey & Wandelt 2020)

$$\mathcal{L} = \left( \sum_{i \in \text{batch}} (\theta_i - \mu_i)^2 \right) + \left( \sum_{i \in \text{batch}} \left( (\theta_i - \mu_i)^2 - \sigma_i^2 \right)^2 \right), \quad (\text{C1})$$

where  $\theta_i$  is the true simulation parameter of the  $i^{\text{th}}$  simulation in the batch used to train the model and  $\mu_i$  ( $\sigma_i$ ) are the mean (standard deviation) of the posterior given as model outputs during training. The number of epochs is tuned from 300 to 1000 epochs, over which there were no significant effects on the results.

The hyperparameters of the model are the number of fully connected layers (NL), the number of nodes in each fully connected layer (NN), the learning rate (LR), the weight decay (WD), and the dropout rate (DR). The LR determines how quickly the neurons within the machine learning model can change during training, the WD regularizes the network by imposing a penalty for large neuron weights, and the DR is the fraction of connections between neurons that are not used in a given epoch during training. These parameters are varied within these ranges:

$$\begin{aligned} \text{NL} &\in [1, 5] \\ \text{NN} &\in [4, 10^3] \\ \text{LR} &\in [10^{-5}, 10^{-1}] \\ \text{WD} &\in [10^{-8}, 1] \\ \text{DR} &\in [0.2, 0.8]. \end{aligned}$$



The LR, WD, and DR are all sampled logarithmically during hyperparameter tuning. A unique value of NN is chosen for each layer within the given network. Hyperparameter tuning is done with OPTUNA (Akiba et al. 2019) over 50 trials.

The standard deviation of the posterior provides the aleatoric error on the result. To include the epistemic error, we average the results of an ensemble of trained models. For the results presented in Section 3.2, we independently train 10 instances of the emulator with the same tuned hyperparameters but different initialization of weights. The mean values presented correspond to the mean result across all ten instances of the emulator. The aleatoric uncertainty is taken as the average width of the posterior from the ten instances of the network. The epistemic uncertainty is the standard deviation of the mean of the posterior across all ten instances of the emulator. The uncertainty that we report is the aleatoric uncertainty added in quadrature to the epistemic uncertainty.

For all results presented in Section 3.2, the aleatoric uncertainty is greater than the epistemic uncertainty. On average, the aleatoric uncertainty is  $\sim 7 \times$  greater than the epistemic uncertainty. However, there is a trend with WDM particle mass where larger masses result in a larger epistemic uncertainty, but still less than the aleatoric uncertainty.

The right panel of Fig. 3 compares the output of the emulator with the simulation data to validate the results. The emulator data (black) represents the expected number of satellites and the uncertainty from both cosmic variance and astrophysical variations. The simulation data (orange) represents the average number and standard deviation of satellites for all simulations binned into 0.5 keV bins. We include astrophysics variations, instead of limiting them to the fiducial values, to have a consistent comparison with the entire simulation dataset.

At low and moderate WDM masses ( $M_{\text{WDM}} \lesssim 12$  keV), the emulator and simulation results agree well. At high masses, the simulation data becomes more noisy and oscillates around the average emulator output due to the lower number of simulations with large WDM mass. There are only 95 simulations with a WDM particle mass greater than 12 keV and only 32 greater than 20 keV. While these results confirm the accuracy of the emulator, we suggest caution in overinterpreting differences between models with WDM particle masses greater than 12 keV.

### C.2. Convolutional Neural Network

Section 4 presents a first application of the hydrodynamic WDM uniform-box suite, using a convolutional neural network (CNN) to predict the WDM particle mass from astrophysical images. The neural network takes as input 2D projected images of DM and/or baryonic density fields (see Fig. 4 for examples). Using the VOXELIZE code,<sup>9</sup> each image is created from the simulation output by first reconstructing the 3D density grid, with a resolution of  $(10 h^{-1} \text{ kpc})^3$ , for a property of interest. The algorithm used to recreate the density field first computes the distance from each simulation particle to its 32 nearest neighbors. The value of 32 affects the smoothness of the reconstructed density field but is commonly used during this kind of reconstruction (eg. Hernquist & Katz 1989; Monaghan 1992). Each particle is then represented by a uniform sphere whose density (associated with the used quantity) is deposited into a regular grid (for details, see Villaescusa-Navarro et al. 2022b).

After generating the 3D density grid, we produce the 2D images by taking  $5 h^{-1}$  Mpc-thick slices of the voxelized grid and projecting them along the short axis. The thickness of  $5 h^{-1}$  Mpc is chosen to match that used for the CAMELS project (Villaescusa-Navarro et al. 2021a) and for Rose et al. (2024); it allows for a balance between a larger dataset and more information contained within each image. We produce 15 unique images per simulation by taking 5 slices along each of the 3 axes of the simulation box and then projecting along the short axis of each slice. Each image is then augmented using four  $90^\circ$  rotations and two reflections. Thus, there are 120 images per simulation for a total of 122,880 images for a suite containing 1,024 simulations. Before training, we normalize the dataset by taking the log of all pixels, normalizing each pixel to the average value of the entire dataset, and then setting the standard deviation across all pixels to one. This normalization is not done on an image-by-image basis, so relative differences between different simulations will persist in the images created from those simulations.

We create images for eleven different astrophysical properties to use during training. These properties are:

$M_{\text{TOT}}$	- the total mass density.	$M_{\text{DM}}$	- the DM mass density.
HI	- the neutral hydrogen mass density.	ne	- the electron number density.
$M_{\text{GAS}}$	- the total gas mass density.	T	- gas temperature.
P	- gas pressure.	$V_{\text{DM}}$	- DM velocity modulus (speed).
$M_{\text{STAR}}$	- stellar mass density.	Z	- gas metallicity.
B	- magnetic field modulus.		

<sup>9</sup> <https://github.com/leanderthiele/voxelize>

See Fig. 4 for examples of each property.

The network architecture consists of 19 convolutional layers, each followed by a batch normalization and a leaky ReLU activation function. The overall size of each image is reduced from  $256 \times 256$  pixels to  $128 \times H$  nodes, where  $H$  is a hyperparameter (see below). The resulting nodes are then passed through a fully connected layer, with dropout, reducing the stack to  $64 \times H$  nodes. These nodes are then passed through a leaky ReLU activation function, dropout is applied again, and the nodes are passed through another fully connected layer to produce the simulation outputs.

The training, validation, and testing sets are split at the simulation level so images from one simulation are not used in both training and validation/testing. This ensures that the models see no overlap in the parameter space between the training and testing image sets. We reserve 5% of the simulations (6,144 images) for the testing and validation sets each, then use the rest for the training set. We have verified that reserving larger fractions for the validation and testing sets does not produce any substantial changes to the results. The results that we present in Section 4 come from the testing set to ensure that our neural network has never seen the data used to produce these results.

This architecture takes a single image as input and outputs two values: the mean and standard deviation of the posterior distribution for the desired property of the simulation. The mean and standard deviation of the posterior are calculated through the loss function given in Equation C1. We utilize a batch size of 128 images but find that other values between 32 and 1024 also produce similar results. Training is done over 200 epochs.

During training, we tune four hyperparameters over 50 trials for each model. In addition to the number of hidden layers,  $H$ , see above, the other hyperparameters include the learning rate (LR), dropout rate (DR), and weight decay rate (WD). The LR controls the average amplitude of the changes that are made to the model parameters between epochs. The DR controls what fraction of nodes in the fully connected layer are removed during training. The WD rate controls how quickly the parameters in the fully connected layer decay during the training process. The four hyperparameters are varied within the following ranges:

$$\text{LR} \in [10^{-5}, 5 \times 10^{-3}]$$

$$\text{DR} \in [0, 0.9]$$

$$H \in [6, 12]$$

$$\text{WD} \in [10^{-8}, 10^{-1}]$$

The trained CNN model used in this analysis has the set of hyperparameters that produce the lowest value of the loss function when evaluated on the validation set. For each of our models, we find that the training, validation, and testing losses are similar, which suggests that they do not suffer from overtraining.

To incorporate the uncertainties intrinsic to the machine learning method, we train an ensemble of five CNNs to measure the epistemic uncertainty. Each CNN is retrained using the optimal hyperparameters from our hyperparameter tuning but with different initial parameter weights. We take the aleatoric uncertainty as the average width of the posterior and the epistemic uncertainty as the standard deviation of the mean of the posterior from the five networks divided by the square root of the number of trained CNNs. We then present the total uncertainty as the aleatoric uncertainty added in quadrature to the epistemic uncertainty.