Decoding Dark Matter Halos through the lens of Machine Learning

by

Tri V. Nguyen

B.S., University of Rochester, 2019

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

© 2024 Tri V. Nguyen. This work is licensed under a CC BY-NC-ND 4.0 license.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Tri V. Nguyen
Department of Physics
August 16, 2024

Certified by: Lina Necib
Assistant Professor of Physics, Thesis Supervisor

Accepted by: Lindley Winslow
Associate Department Head
Department of Physics

Decoding Dark Matter Halos through the lens of Machine Learning

by

Tri V. Nguyen

Submitted to the Department of Physics
on August 16, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

Dark matter (DM) constitutes about 85% of the matter in the Universe, yet its particle nature remains one of the greatest outstanding questions in astrophysics. DM halos act as the scaffolding within which galaxies form, but the specific mechanisms through which they influence galaxy evolution are not fully understood, especially at galactic scales. While cosmological simulations and astrophysical surveys have made significant strides in constraining DM properties, upcoming surveys will generate terabytes of complex, high-dimensional data. It is thus imperative to develop new methodologies capable of interpreting and linking this data with theoretical models. Machine learning techniques, coupled with advancements in cosmological simulations, present a transformative opportunity.

In this thesis, I conduct a multi-scale investigation into the nature of DM and its role in shaping galaxies by integrating advanced machine-learning techniques with cutting-edge cosmological simulations. First, I employ simulation-based inference and graph neural networks to infer the mass density profiles of DM halos in dwarf galaxies from their stellar kinematics. Next, I develop a generative model using normalizing flows and recurrent neural networks to reconstruct the mass assembly histories of DM halos in cosmological simulations. Furthermore, I utilize variational diffusion models and Transformer-based neural networks to perform point-cloud modeling of satellite populations under alternative DM models. Finally, I create synthetic surveys for the Gaia surveys from Milky Way-like simulations, bridging the gap between simulations and observations.

This thesis demonstrates the transformative potential of machine learning techniques to probe the DM properties and galaxy formation. The methodologies developed herein provide new avenues for interpreting vast and complex astronomical datasets and offer insights that could lead to a deeper understanding of the fundamental nature of DM.

Thesis supervisor: Lina Necib
Title: Assistant Professor of Physics

Acknowledgments

My Ph.D. journey has been long and arduous, full of ups and downs, unexpected turns, and plot twists. I made it through eventually, but it wasn't without blood, sweat, and tears. And certainly, it wasn't without the help and support of friends and colleagues, many of whom I consider family. Each of you has come into my life at different times and in different ways, but all have left an impact. With this acknowledgment, I dedicate this Ph.D. to all of you. Cheers.

First, I want to thank my awesome advisor, Lina Necib. At the start of my third year, I switched to Lina's group from a completely different field, and wow, what an experience! Thank you for challenging me intellectually and always pushing me to be my best self. For your unconditional kindness, patience, and understanding whenever I make mistakes. For saying yes and for saying no, as I know you always have my best interest in mind. For fighting to make my internship at the CCA happen. For all your time removing outrageous memes from my presentations (and this thesis) and dealing with my many other shenanigans. For introducing me to the Dancing Doge emoji so that I can keep spamming everyone with it. Working with you is the best thing that happens to me at MIT. You saved my career, and I can't ever say thank you enough.

Throughout my time at MIT, I have had many mentoring figures who have taught me so much and endlessly supported me through many stages of my career. In particular, I would like to thank Siddharth Mishra-Sharma and Nora Shipp. Thank you both for many fun science discussion, invaluable career advice, and all the Rumor Mill gossiping. I have learned so much from you, and I can't image going this far without you. I also want to thank my academic advisor, Scott Hughes, for your many words of advice, especially during the early years of my Ph.D. when I was the most vulnerable, and for being the initial bridge between me and Lina.

I would also like to acknowledge the MIT Physics Department and the Astrophysics Division community. Champ, thank you for the memorable board game nights and for showing me So.Many.Restaurants around Boston/Cambridge. Sylvia, thank you for being an amazing friend and office mate and letting me crash at your apartment and pet Charlotte

and Andy. Stephanie and Lisa, thank you for all the good memories, endless snacks, and the patience to deal with my shenanigans. Thank you Alex, Cian, Yadira, Sydney, Xiaowei, Ellie, Michael, Danielle, Katiya, Joheen, Mason, Teo, Swati, Noah, Ethan, Ken, Caolan, Sam, and Wenzer for being the best and unforgettable parts of MIT. I would also like to thank Cathy, Sydney, and Deepto for all your support, Nayanika for ordering the best possible food for Journal Club, Ty'Shauna for organizing my thesis defense's celebration party, and Thea for helping me process my reimbursement and stay financially stable.

My grad school life would not be nearly as memorable and fun without the support and fellowship of many friends in the Boston/Cambridge area. I would like to thank Phuc and Thy for showing me the kinda-meh food scene around Boston/Cambridge when I was still new to the city. I hope our chopsticks and tapioca will cross again one day in the near future. Buu and Diem, you guys are the best two wheels that this third wheel could have ever asked for. I will always treasure all our adventures and moments we spent together, and I wish you all the happiness in the world. Thank you, Phuc (Tung), for showing me around Providence every now and then and for many fun and questionable conversations we've had. Thank you, Trung, for occasionally visiting and for being the only other Vietnamese I know in astrophysics and CCA. I would also like to acknowledge my friends in the Cambridge/Boston breaking community, in particular the TFC (The Flavor Continues) and MIT Imobilare. Break dancing (or breaking) takes my mind off of things and helps me stay sane during grad school. Thank you Brian, Jamie, Nha, Kaleb, and Michelle for organizing many practice sessions and jams and for helping me get my Windmill. I wish that the The Flavor Continues's flavor continues forever.

The Center for Computational Astrophysics (CCA) has been an unforgettable part of my Ph.D. Starting in September 2022, I came to the CCA for a semester as a Pre-doctoral Researcher. It was not an overstatement to say that this experience was the best 5 months of my life. I love everything: the amazing free food, the vibrant academic community, the amazing free food, the unlimited computational resources, the sleeplessness of New York City, and the amazing free food. Even after the semester was over, the CCA continued being a major player in my academic career, hosting many key conferences and workshops and providing me still with their seemingly unlimited computational resources. Thus, I would like to acknowledge the CCA and its community for giving me this amazing, life-changing opportunity. I would like to thank my advisors, Rachel Somerville, Aaron Yung, and Chirag Modi for providing me with your endless support and patience and challenging me every day. I'm incredibly excited to see the science we will do with FLORAH!

I have met many amazing friends and colleagues at the CCA, whom I will never forget. I would like to thank my fellow beloved pre-doctoral friends, Alba, Arjun, Chris, Jonas, and

Ruby (honorary). I love you guys from the very bottom of my heart, and I wish you all the happiness and wisdom of the Universe. Chris and Jonas, thank you for being the best housemates I could've ever asked for and the best human beings this planet could've ever asked for. Ka-Kaw! Alba, thank you for being my partner-in-crime and enabling my many shenanigans. Arjun, your happiness is always contagious, and thank you(?) for introducing me to BeReal. Ruby, thank you for the life-saving water boiler and all the yee-haws. I would also like to thank Sandip, Dylan, and Adri for being amazing friends and colleagues. I enjoy all of our dark matter discussions and thank you for keeping me curious! Last but not least, I would like to thank Adri for hosting my job talk at Princeton and telling Lina that I identify myself as Chaotic Neutral.

Over the years, I am also incredibly fortunate to have had many awesome collaborators, who have challenged and supported me intellectually and shown immense interest in my work and career. I want to thank my collaborators and colleagues, Francisco Villaescusa-Navarro, Jonah Rose, Paul Torrey, Mariangela Lisanti, Scott Dodelson, Natalí Soler Matubaro de Santi, Justin Read, and Carolina Cuesta-Lazaro, I am excited to see what we will achieve together in the next few years. I would also like to thank my incredible students, Chelsea, Rutong, and Zeineb, who have significantly enriched my Ph.D. experience.

Lastly, rewinding to the olden days at the University of Rochester, to a time of youth, exploration, naivety, and *terrible, terrible, terrible food*, I would like to acknowledge Eric Mamajek, who taught my intro astronomy class and encouraged me to pursue astronomy as a career path. Without you, I would have switched to Optical Engineering, where the pay is higher but the happiness is lower and the soccer is worse.[1] I would also like to thank Regina Demina and Segev BenZvi, who were my undergraduate research advisors, for all their patience, support, and advice. Regina, I have learned so much about particle physics and cosmology from you. Segev, the Python and Bayesian statistics that you taught me are undoubtedly the foundation for my machine learning. I would also like to thank Michael Coughlin, my Caltech SURF advisor, who introduced me to machine learning and unrelenting support throughout the first two years of grad school. I also want to take a moment to acknowledge URochester's Physics and Astronomy student community. Thank you, Ryan, Bo, Sanha, and Tanveer for being such good sports and for your many words of senior advice! Sam, Will, Jeremy, and Tony, you guys were an unforgettable part of my social life! Thank you Pouya for all your fun ~~tortu~~training sessions at the gym! I felt I became more resilient and disciplined after every session. Last but not least, Genevieve, thank you for being an awesome friend and hosting my job talk at CIERA!

[1] The final scores of the annual Physics/Astro vs Optics soccer game in 2015: 7-1. Take that Optics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Dark Matter Halos and their Galaxies

Dark Matter (DM) is a form of matter that does not interact with electromagnetic waves like light [1]. Unlike normal matter, it cannot absorb, emit, or scatter light and is thus invisible to conventional detectors such as telescopes. Though we have not yet detected DM, we have strong evidence for its existence through its gravitational influence on its surroundings, e.g., dynamical observations of stars and galaxies, gravitational lensing [1], [2]. Because DM constitutes about 85% of matter in the Universe, it has important implications for the dynamics and evolution of stars, galaxies, and even the Universe itself [3]. Still, the nature of DM remains one of the greatest outstanding questions in modern astrophysics and has been the priority of many scientific experiments since the 20th century [4], [5]. The discovery of DM is bound to be a major breakthrough that revolutionizes physics and astronomy.

1.1.1 Observational evidence for Dark Matter

In this section, I discuss the key observational evidence that supports the existence of DM:

1. **Galactic Rotation Curves**: A *rotation curve* of a galaxy describes how the orbital velocity of its stars changes as a function of distance to the center of the galaxy. The observed rotational velocities are related to the enclosed mass via the relation:

$$v(r) \propto \sqrt{\frac{GM(r)}{r}} \tag{1.1}$$

   where G is the gravitational constant, r is the radius to the center of the galaxy, $v(r)$ is the rotational velocity, and $M(r)$ is the enclosed mass. By measuring the orbital velocity, which can be obtained from spectroscopic observations of the Doppler

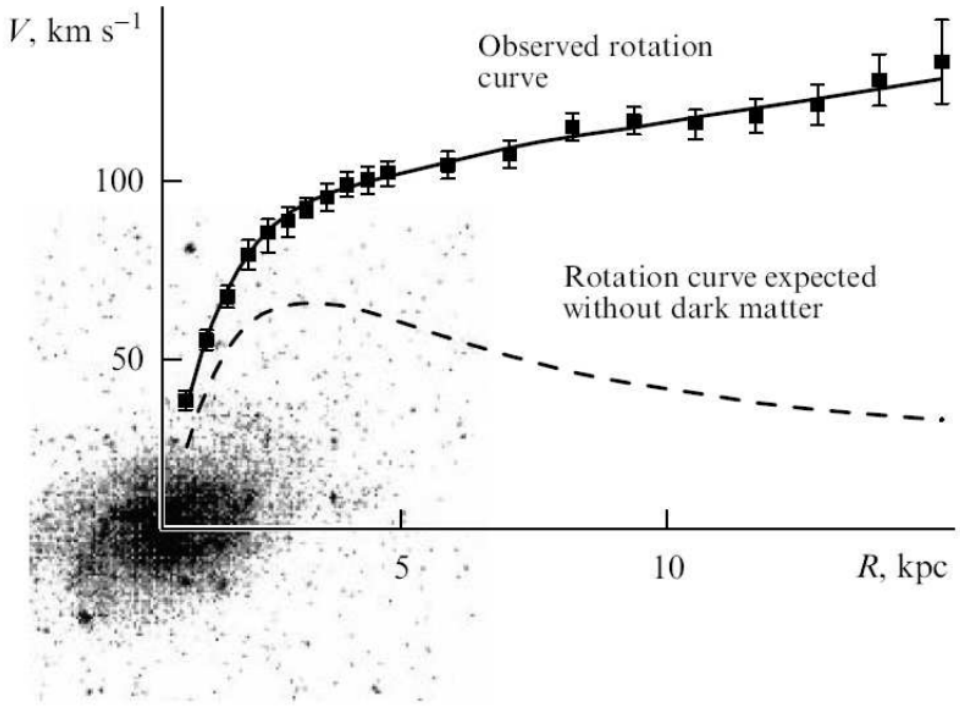

Figure 1.1: The rotation curve of the Triangulum Galaxy (M33) suggests the galaxy is dominated by an invisible mass component that extends beyond the optical radius of the galaxy. Figure adapted from [11].

shift of stellar or gas emission lines (see Ref. [6] for a review), we can infer the mass distribution of both visible and invisible matter. Observations of the rotation curves of galaxies show that the rotational velocity of stars remains constant or even increases with distance from the galactic center [7], contrary to what would be expected if only visible matter were present. This discrepancy implies the presence of additional, unseen mass (e.g., [7]–[10]).

Figure 1.1, adapted from [11], illustrates how the rotational velocity of the Triangulum Galaxy (M33) increases with increasing distance from the galactic center. Based on the amount of visible mass presented in M33, we would expect the rotational velocity to be decreasing with distance, following a Keplerian decline. Instead, the observed increasing rotation curve indicates the presence of a dominant invisible mass component, which we then call DM, that extends well beyond the optical radius of the galaxy.

2. **Dynamics within Clusters**: The dynamics of stars within dwarf galaxies and galaxies within galaxy clusters further support the existence of DM. In particular, the velocity dispersion σ, which is a measure of the range of velocities of stars or galaxies in

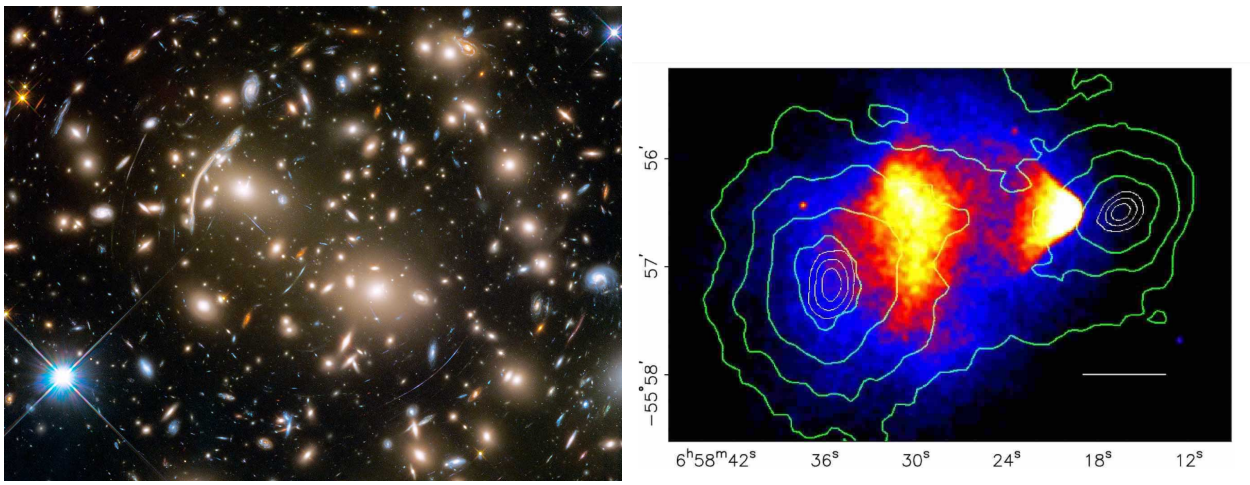

Figure 1.2: *Left*: An image of the Abell 370 galaxy cluster taken by the Hubble Space Telescope, showing several hundred galaxies. The arcs of blue light are distorted images of galaxies located behind the cluster, caused by the gravitational lensing effect. Credit: NASA, ESA, and J. Lotz and the HFF Team (STScI). *Right*: An X-ray image of the Bullet Cluster taken by the Chandra X-ray Observatory, adopted from [13]. The contours show the mass distribution, which is offset from the distribution of hot X-ray gas.

a system, scales with the enclosed mass:

$$\sigma^2 \propto \frac{GM(r)}{r} \tag{1.2}$$

Observations of galaxy clusters show that the velocity dispersions observed in these systems are too high to be explained by visible mass alone [12]–[14]. Likewise, studies of dwarf galaxies, which have relatively little visible matter, also show that their dynamics and velocity dispersion can only be explained if a substantial amount of DM is present (e.g., [15]–[18]).

3. **Gravitational Lensing**: Predicted by Einstein's General Theory of Relativity, gravitational lensing is a phenomenon in which the gravitational field of a massive object, such as galaxies and galaxy clusters, bends the light from an object behind it [19], [20] (see Figure 1.2). This effect allows astrophysicists to detect and map the distribution of mass that is difficult or impossible to observe through electromagnetic radiation, including DM, black holes, faint stars, and even extrasolar planets. Observations of gravitational lensing in galaxy clusters and individual galaxies reveal more mass than can be accounted for by visible matter alone, providing strong evidence for the presence of DM (see Ref. [21] for a review).

4. **Bullet Cluster**: The Bullet Cluster (1E 0657-56) is a compelling example of gravi-

tational lensing that provides direct evidence for DM [13]. It consists of two galaxy clusters that have collided, with the hot gas from each cluster interacting and slowing down, while the galaxies and DM passed through each other with minimal interaction.

Lensing maps of the Bullet Cluster reveal two distinct mass concentrations that are spatially separated from the hot gas (see Figure 1.2). These indicate that most of the mass is associated with the DM and galaxies, rather than the hot gas. This separation also demonstrates that DM behaves differently from normal matter, as it might not significantly interact with itself [22] or the hot gas during the collision [13], [23]. Furthermore, this observation challenges modified gravity theories [24], which struggle to explain the distinct separation of mass [25], [26].

5. **Cosmic Microwave Background**: The Cosmic Microwave Background (CMB) is the afterglow of the Big Bang and contains small temperature fluctuations that reflect the density variations in the early Universe. The scales of the density fluctuations, commonly called the anisotropies, are sensitive to models of DM and cosmological parameters. Observations from missions such as COBE (Cosmic Background Explorer, [27]), WMAP (Wilkinson Microwave Anisotropy Probe, [28]), and Planck [3] have measured these anisotropies with high precision. The power spectrum of the CMB anisotropies, particularly the first three acoustic peaks, aligns with models that include DM [28]–[32]. These observations indicate that DM constitutes about 85% of the total matter content in the universe, influencing the formation and evolution of large-scale structures.

6. **Large-scale Structure**: The distribution and evolution of galaxies and galaxy clusters on cosmic scales provide additional evidence for DM. Simulations of large-scale structure formation require DM to reproduce the observed filaments structure of the Universe. These simulations match the distribution of galaxies observed in large surveys, such as the Sloan Digital Sky Survey (SDSS, [33]), only when DM is included [34].

These various lines of evidence for the existence of DM, all originating from astrophysical observations, paint a compelling picture of a universe dominated by an invisible form of matter. These astrophysical observations not only suggest the presence of DM but also hint at its critical role in shaping the cosmos. To further understand the implications of DM, we must understand the theoretical frameworks and simulations that describe the formation and evolution of the so-called *dark matter halos*.

1.1.2 Connection between Halos, Galaxies, and Stars

DM particles, due to their gravitational interactions and negligible self-interactions [22], [35]–[38], tend to collapse into highly dense and roughly spherical regions known as *dark matter halos*. According to the standard cosmological model, ΛCDM (Lambda Cold Dark Matter), these halos form through a hierarchical process [39]. In the early Universe, small DM halos merged and accreted matter, eventually forming larger and more massive halos over time.

DM halos play a crucial role in the formation and evolution of baryonic structures. Within these halos, baryonic matter, primarily gas, cools and collapses under gravity, leading to the formation of stars and galaxies. Consequently, galaxies are enclosed within DM halos. Since DM halos typically account for the majority of mass in galaxies, they have significant influences on the dynamics, morphology, and evolution of the stars and gas within galaxies. *Therefore, understanding the properties and distribution of DM is essential for comprehending galaxy formation and evolution.* The interplay between DM and baryonic matter dictates various processes such as star formation rates, feedback mechanisms from supernovae and active galactic nuclei, and the overall structural evolution of galaxies. High-resolution cosmological simulations incorporating both DM and baryonic physics, such as the IllustrisTNG [40], [41], EAGLE [42], FIRE [43] simulations, have provided insights into these complex interactions. By studying these simulations alongside observational data, we can test and refine models of galaxy evolution within the framework of ΛCDM.

1.1.3 Constraining Dark Matter with Galaxies and Stars

In the previous section, we discuss how insights into the DM models can help understand the physical processes behind galaxy formation and evolution. Conversely, galaxies and their stellar populations can be used as tracers to study the distribution of DM in galaxies and clusters, which is sensitive to the underlying DM model (see [44], [45] for a review). For example, cold dark matter (CDM) models predict numerous small halos and cuspy density profiles due to the absence of significant DM particle interactions. Warm dark matter (WDM) models, with slightly more energetic particles, predict fewer small-scale structures and smoother DM distributions, as these small-scale fluctuations are suppressed (e.g., [46], [47]). Self-interacting dark matter (SIDM) models introduce interactions between DM particles, leading to core-like density profiles in the centers of halos due to scattering (e.g., [36], [48], [49]).

In this section, I highlight a few ways galaxies and stars can help differentiate between DM models, starting with observations within the Milky Way:

- The Gaia mission [50] creates a precise 6D map of the positions and velocities of billions

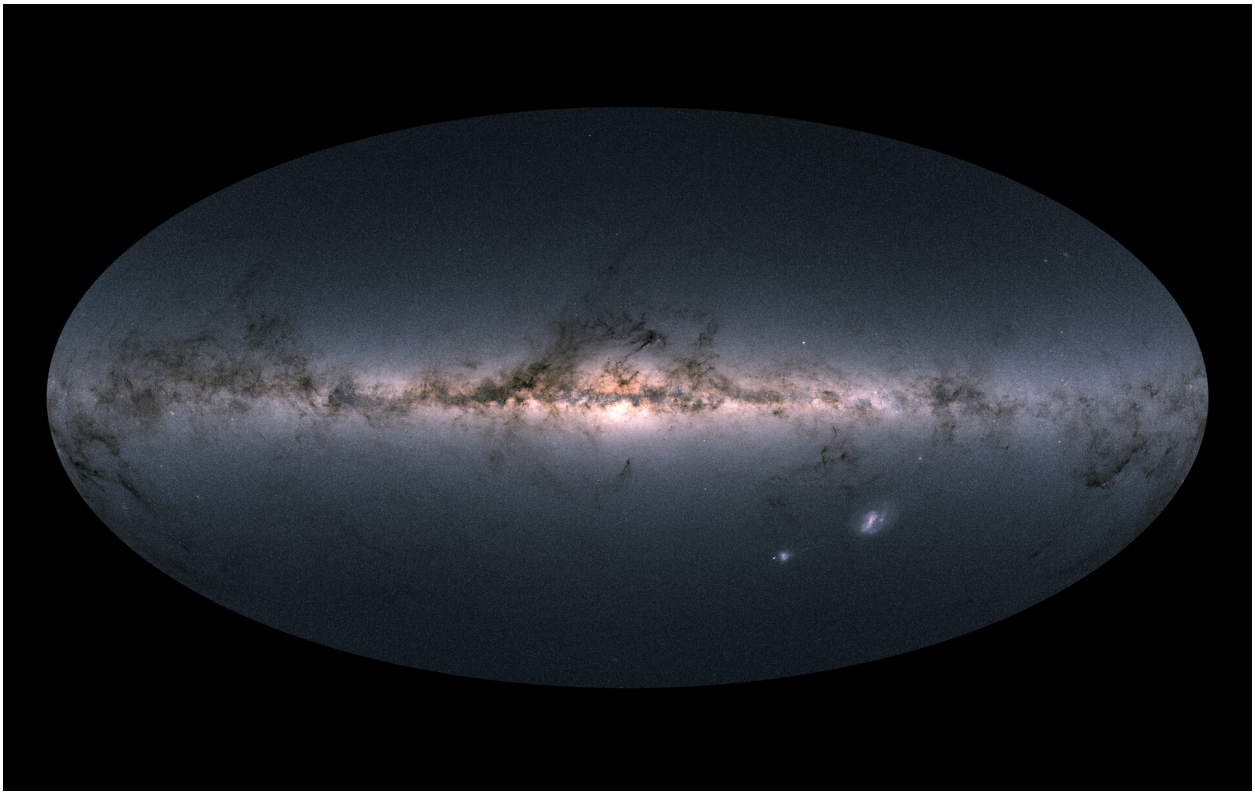

Figure 1.3: An image of the Milky Way by the Gaia space telescope showing billions of stars. Credit: ESA/Gaia/DPAC, CC BY-SA 3.0 IGO.

of stars in the Milky Way (see Figure 1.3). Stellar motions provide critical data for modeling the Milky Way's gravitational potential and constructing its rotation curve, which is essential for understanding the distribution of both visible and dark matter in our galaxy. By accurately mapping these stellar motions, Gaia helps constrain the distribution and density of DM within the Milky Way and subsequently informs the DM models (e.g., [51]–[53]).

- Stellar motions can also be used to trace the local DM distribution near the Solar neighborhood [54]. This has direct implications for DM direct detection experiments, such as XENONnT [55], LZ (LUX-ZEPLIN, [56]), etc. These experiments search for interactions between DM particles and atomic nuclei in the detector materials and rely on accurate models of the DM local velocity distribution and density to predict the flux and scattering rates of DM particles.

- Galaxies undergoing mergers with the Milky Way form elongated stellar structures known as "stellar streams" (see Figure 1.4). When a stellar stream passes through a DM subhalo, it experiences disruptions that can be detected as gaps or shifts in

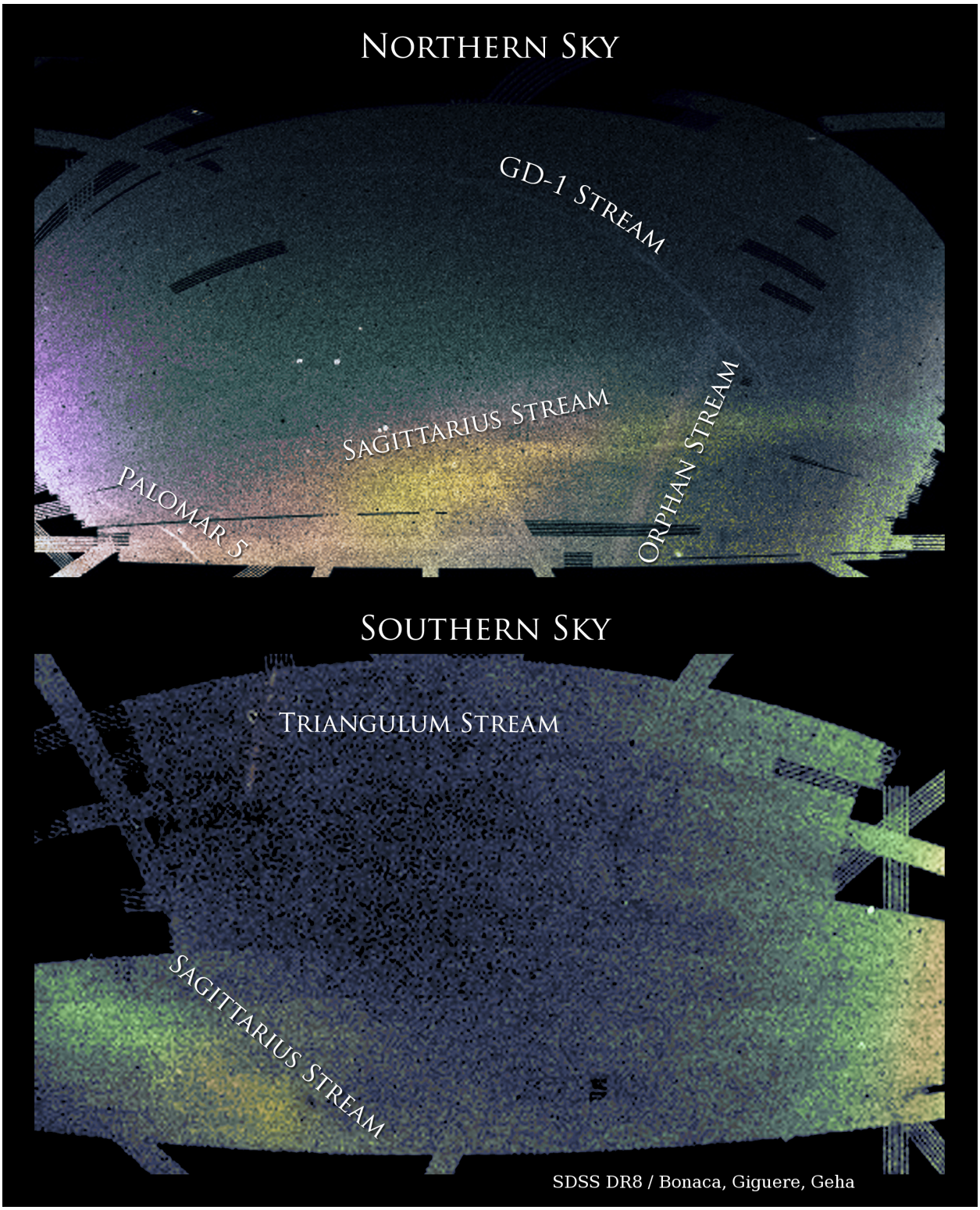

Figure 1.4: Density mapping of the Northern and Southern Galactic Hemisphere from SDSS showing various elongated stellar structures known as stellar streams. These streams are remnants of galaxies and star clusters that are accreted onto the Milky Way. Credit: SDSS DR8 / Bonaca, Giuere, Geha.

the stream [57]–[60]. These DM subhalos are often not massive enough to host stars (known as dark subhalo) and thus are challenging to detect. Mapping stellar streams with surveys like Gaia, S5 (Southern Stellar Stream Spectroscopic Survey [57], [61]), DES (Dark Energy Survey [60], [62]), can help detect and constrain the abundances and density profiles of these subhalos, which are sensitive to the properties of DM [57], [58]. For example, the presence and frequency of these disruptions can differentiate between CDM, which predicts numerous small subhalos, and WDM, which predicts fewer subhalos [63].

- Spectroscopic surveys such as SDSS (Sloan Digital Sky Survey [33]) and APOGEE (Apache Point Observatory Galactic Evolution Experiment [64]), can help constrain the properties of DM. The stellar spectra measured by these surveys are used to determine the radial velocities of stars more accurately, as well as their chemical compositions. By studying the motions and chemical compositions of stars, we can identify past merger events and accretion of smaller galaxies (e.g., [59], [65]–[69]). These merger events can be used to understand the hierarchical formation of the Milky Way, which is a key prediction of the ΛCDM model. Differences in the predicted and observed merger histories (e.g., accretion rate, formation time, etc.) can indicate the need for modifications in the DM model and even cosmology [63], [70], [71].

- Detailed chemical abundances provided by spectroscopic surveys are required to accurately model baryonic processes such as star formation histories and the impact of supernovae and stellar winds on the interstellar medium. Star formation and feedback processes can alter the distribution of DM in galaxies in a similar way to the effects predicted by alternative DM models [72]–[75]. For example, SIDM models predict core-like density profiles due to DM particle interactions, which can also result from baryonic feedback processes like supernovae [76]–[78]. The degeneracy between baryonic feedback effects and those predicted by alternative DM models is among the most significant challenges in understanding DM properties [79]. By accurately modeling these baryonic processes, we can differentiate between the effects caused by baryonic feedback and those intrinsic to DM models, breaking this degeneracy.

In addition to the Milky Way, observations of *dwarf galaxies* can also help constrain DM models. Dwarf galaxies are small, DM-dominated galaxies that are often found orbiting larger, Milky Way-mass halos [80]–[83] (see Figure 1.5). They are among the first galaxies to form and thus encode the density conditions of the early Universe. Moreover, due to their lack of baryonic matters (e.g., stars and gas) relative to Milky Way-mass galaxies, they are perfect testbeds of DM and have provided some of the strongest constraints on DM

properties (see Ref. [45], [83] for a review). Indeed, as we currently lack a comprehensive understanding of the small-scale effects of DM and their implications for galaxy formation, dwarf galaxies are among the most promising avenues for testing DM models in the near future. Below, I discuss a few key observations of dwarf galaxies and their implications.

- The inner density profiles of dwarf galaxies are sensitive to the underlying DM models. CDM predicts steeply rising density profiles (known as cuspy profiles), while alternative DM models such as WDM and SIDM can predict constant density cores (known as cored profiles). Unlike Milky Way-mass galaxies, dwarf galaxies, especially ultra-faint dwarfs, lack significant stars and gas, making them less affected by baryonic processes that can alter their DM profiles [85], [86]. This provides near-pristine conditions for testing DM models. To measure the inner density profiles and infer DM properties, we can analyze the stellar dynamics in these galaxies, using radial velocities data from high-latitude spectroscopic surveys such as APOGEE and DES (e.g., [87]–[93]).

- The inner density profiles also have implications for DM indirect detection experiments, which search for signatures of DM annihilation and decay. DM particles can annihilate or decay into standard model particles, producing high-energy photons such as γ-rays. These γ-rays can be detected by telescopes such as Fermi-LAT [88], [94]–[101]. The rate of DM annihilation or decay is proportional to the square of the DM density (for annihilation) or linearly proportional to the DM density (for decay) [102], making the inner density profiles of dwarf galaxies critical for these searches. Since dwarf galaxies, especially ultra-faint dwarfs, have high DM densities and low astrophysical backgrounds, they provide ideal environments for indirect detection experiments [88], [94]–[101].

- Other properties of dwarf galaxies, such as their abundances and orbits around their hosts, can be used to constrain the DM models. Satellite orbits can be used to reconstruct the Milky Way's gravitational potential and merger histories, which, as discussed, are sensitive to the DM models. The abundance of satellite dwarf galaxies is a key observation for distinguishing between CDM, WDM, and other more exotic models of DM. In the early Universe, WDM particles are more energetic, leading to a free-streaming effect [46], [103] that suppresses structure formation at small scales. In other words, WDM models predict fewer low-mass galaxies within this mass range [104]. This affects the abundance and stellar mass distributions and thus can be used to test the validity of these models. Key surveys such as ELVES (Exploring the Local Volume in the Extended Solar neighborhood, [105]) and SAGA (Satellites Around Galactic

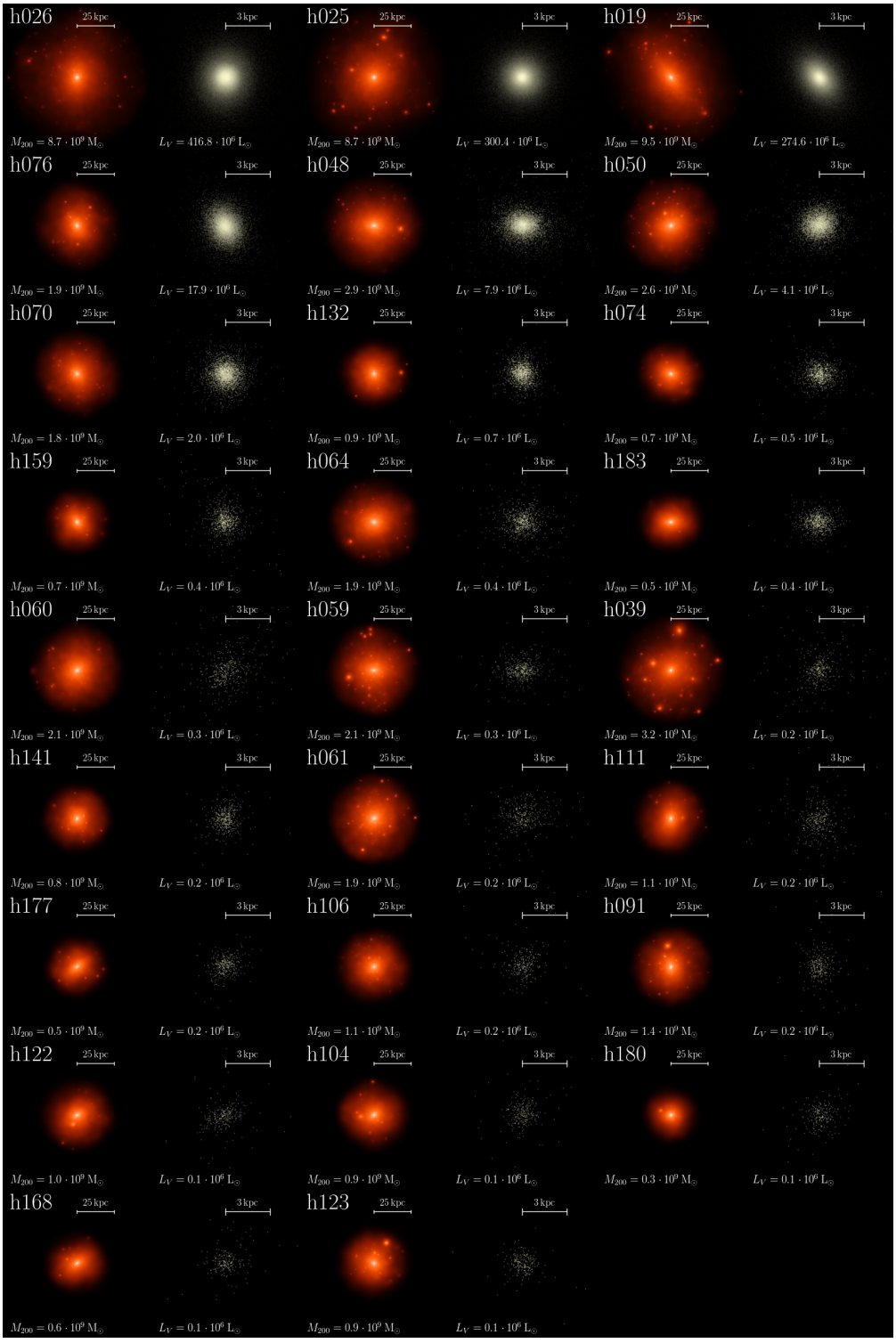

Figure 1.5: A catalog of 27 simulated dwarf galaxies, adopted from [84].

Analogs, [80]–[82]) characterize hundreds of satellite galaxies around nearby galaxies in the Local Volume and beyond.

- As discussed previously, stellar streams and other tidal remnants of accreted dwarf galaxies can be used to detect DM substructure within the Milky Way. Properties of these streams, such as orbital motions, shapes, and chemical compositions, can also be used to reconstruct their galaxy progenitors and past merger events (see [106] for a review). For example, by tracing the path of the stellar stream in the sky and measuring the velocities of its stars, we can reconstruct the orbit of the progenitor. The width and density of a stellar stream provide clues about the mass of the progenitor [57], which can subsequently be used for calculating the subhalo mass functions. The age distribution of stars within a stellar stream can inform the star formation history of the progenitor and the merger time, constraining the models of hierarchical formation.

Observations of the Milky Way and dwarf galaxies in Local Group and beyond have enabled some of the strongest constraints on the particle properties of DM and its role in structure formation at the smallest scales. It is also worth mentioning a few large-scale observations that can help constrain DM.

- For example, large-scale surveys like SDSS, DES, and DESI (Dark Energy Spectroscopic Instrument [107]) map the distribution of millions of galaxies. By analyzing galaxy clustering and gravitational lensing, these surveys help trace the distribution of DM on large scales, which can be used to test predictions of different DM and cosmological models (e.g. [108], [109]).

- Additionally, in recent years, the James Webb Space Telescope (JWST [110]) has delivered unprecedented observations of the early Universe. These observations are not only crucial for understanding the formation and evolution of supermassive black holes and galaxies but also highlight some potential tensions with the current cosmological and DM models. For example, JWST observes an overabundance of galaxies in the early Universe, implying that these galaxies *might* have formed much earlier than predicted by ΛCDM [111], [112]. This suggests that our understanding of galaxy formation, cosmology, and DM is incomplete and opens up exciting avenues for using these early galaxies as probes for DM models.

1.2 The Galactic Frontier of Dark Matter Research in Astrophysics

In the previous section, I discuss the observational evidence for DM and the connection between DM halos, galaxies, and stars. This interconnected relationship implies that to accurately model the formation and evolution of galaxies, we must first understand the nature and behavior of DM. Conversely, to constrain the particle properties of DM, we must study the galaxies and structures it helps form. The study of DM and galaxies is thus deeply intertwined, with each providing critical insights into the other.

My thesis focuses on developing novel techniques that will assist future probes for the nature of DM, using observations of galactic cosmic structures such as the Milky Way and dwarf galaxies. This is among the most promising avenues for DM research and will be the priority for future surveys and telescopes such as the Vera C. Rubin Observatory [113] and the Nancy Grace Roman Space Telescope. Indeed, the role of DM in the formation and evolution of galaxies at small scales is not fully understood. While ΛCDM has many successes in explaining large-scale structures such as the power spectrum of the CMB anisotropy [28], [32] and the distribution of galaxy clusters [114], [115], discrepancies arise at the small scales (see Refs.[79], [116] for a review).

In this section, I will first discuss these small-scale challenges to ΛCDM. Then, I will discuss theoretical and analysis tools, including simulations, synthetic observations, and machine learning techniques, that are being developed to address these challenges and enhance our understanding of DM at small scales.

1.2.1 Small-scale challenges to Lambda-CDM

Cosmological simulations of ΛCDM at the scales of the Milky Way and dwarf galaxies often struggle to reconcile with observations. These challenges point to gaps in our understanding of DM as well as the baryonic processes that govern galaxy formation and evolution. Below I discuss some key challenges:

1. **Core-Cusp**: As briefly discussed previously, DM halos in CDM simulations have density profiles with steeply rising slopes at lower radii (cuspy profiles), while in alternative DM simulations like SIDM, these halos tend to have flat central density profiles (cored profiles). However, recent observations of dwarf and low surface brightness galaxies often show cored DM profiles [117]–[119], contradicting CDM predictions. A potential solution is that baryonic feedback could flatten the density profiles, as has been observed in Milky Way-mass galaxies and galaxy clusters [72]–[75]. However, dwarf

galaxies, especially ultra-faint dwarfs, lack significant baryonic matter and thus *might* not experience substantial feedback [120], [121]. Thus, alternative DM models like SIDM, which can naturally produce cored profiles without the need for strong baryonic processes, *could* provide a better explanation for this core-cusp problem.

2. **Diversity of Shapes**: Observations of the rotation curves of dwarf galaxies show a wide diversity in shape [122]–[125], suggesting a corresponding diversity in their DM halo density profiles. Such variation is not easily reproduced by ΛCDM simulations, which typically predict a more uniform set of rotation curves for dwarf galaxies. In addition, recent studies highlight these discrepancies, showing that some dwarf galaxies have flat, cored profiles while others have steeper, cuspy profiles. The diversity in observed rotation curves indicates that there might be complex interactions between DM and baryonic matter that are not fully captured in current models.

3. **Planes of Satellites**: Observations show that the satellites of the Milky Way and Andromeda often lie in thin, planar structures [126]–[130]. This planar arrangement is difficult to explain under the current ΛCDM paradigm, which predicts a more isotropic distribution. This suggests either a need for new physics or a better understanding of satellite galaxy formation and dynamics. Note, however, that is also unclear if alternative DM models will allow for similar planar configurations of satellites. It is also unknown if the Milky Way and Andromeda are outliers. In the future, surveys such as SAGA and Rubin Observatory's Legacy Survey of Space and Time (LSST) will observe more satellites and can help shed light on this problem [80]–[82], [131].

4. **Too-Big-to-Fail**: ΛCDM simulations predict the existence of several massive, luminous subhalos around the Milky Way. However, these predicted subhalos do not correspond to any observed satellites, which are too small and faint. Specifically, these subhalos appear to be too massive to fail to form stars and become visible satellites, yet they do not appear in the expected numbers or luminosities [132], [133]. This discrepancy challenges our understanding of galaxy formation in massive subhalos and suggests the need for a reassessment of either the DM properties, the baryonic processes, or the detectability limits of current surveys [134], [135]. Future surveys such as Rubin Observatory's LSST are expected to help resolve this problem by increasing the completeness of nearby dwarf galaxies [131].

5. **Missing Satellite**: The missing satellite problem refers to the discrepancy between the number of small satellite galaxies predicted by CDM simulations and the number observed around the Milky Way and other galaxies [16], [136]. This suggests a stronger

baryonic feedback model that suppresses star formation in these satellites or alternative DM models such as WDM that predict fewer low-mass subhalos [63]. However, this problem has been partially resolved due to a better understanding of survey selection effects and deeper observational surveys [137], [138]. Next-generation surveys, such as the Rubin Observatory's LSST, are expected to further improve our understanding by uncovering more faint satellite galaxies [131].

Many of these challenges stem from an incomplete understanding of baryonic processes, including, for example, star formation, supernovae and stellar wind feedback, and the cooling and heating of interstellar gas. These processes are among the key systematic uncertainties in cosmological simulations [121], [139]–[145]. Environmental effects, such as tidal interactions with the Milky Way and other galaxies [132], [146], and observational systematics can further complicate the comparisons between simulations and observational data. While some like the Missing Satellite Problem have partially been resolved, many open questions still remain [147], [148]. This not only motivates alternative theories of DM but also highlights the importance of accurate models for baryonic processes.

In the following sections, I will discuss the tools in modern astrophysics that can address these challenges, including advanced simulations, synthetic surveys, and machine learning techniques.

1.2.2 Cosmological simulations

In the previous sections, I have many times invoked "cosmological simulations." But what are these simulations, and why are they so critical to our understanding of the Universe? Cosmological simulations are sophisticated computational models that evolve the Universe from the initial density perturbations after the Big Bang to the large-scale structures we observe today. These simulations use numerical methods to solve the complex equations governing the interactions of DM, dark energy, and baryonic matter.

The literature surrounding cosmological simulations, including numerical techniques, limitations, and scientific predictions, is comprehensive and well-documented. For a review, refer to Ref. [149] and references therein. Here, I highlight a few key points about these simulations, specifically their techniques, current limitations, and their application to small-scale structures.

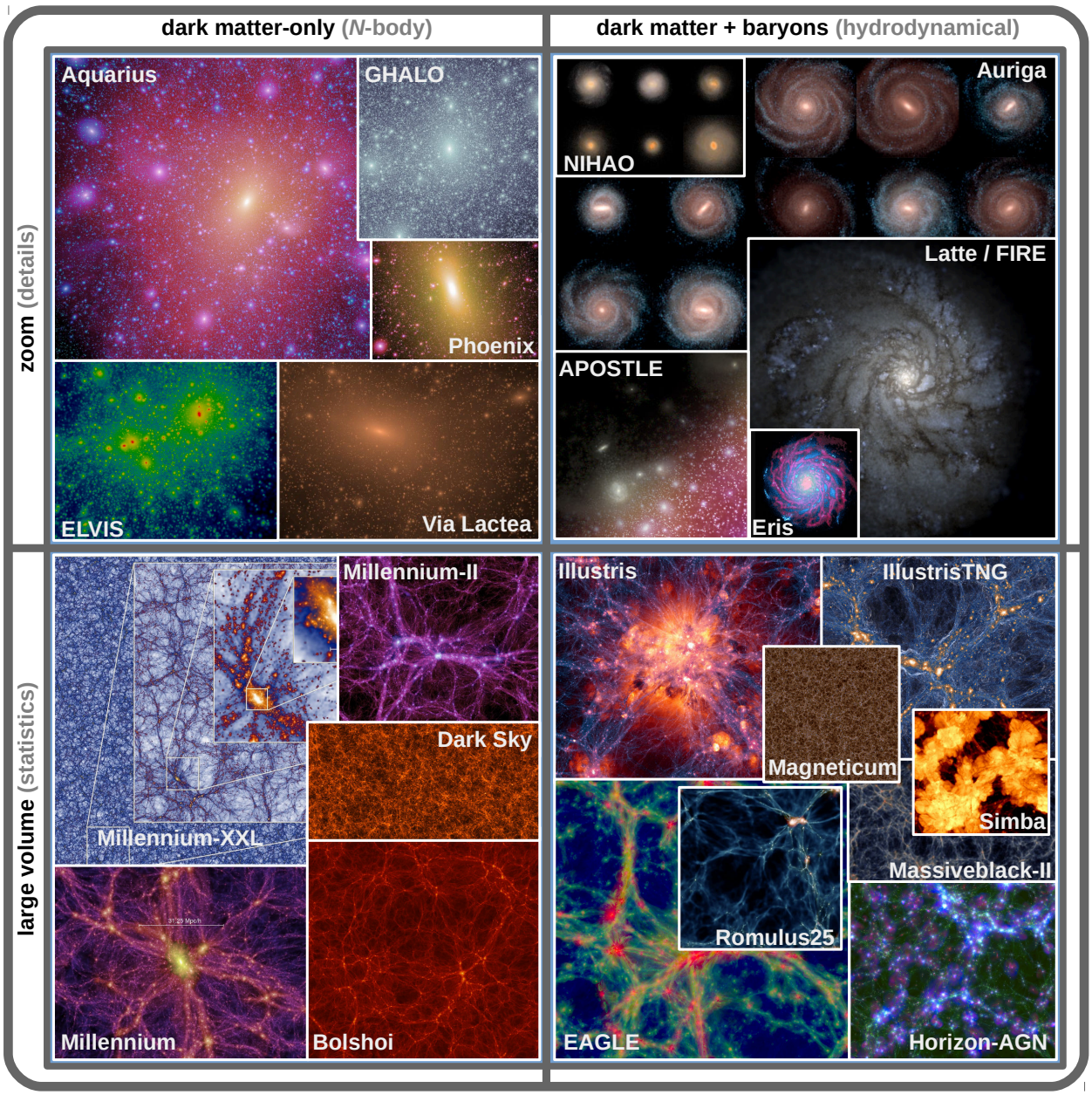

Figure 1.6: Cosmological simulations. Figure adopted from [149].

Overview of Techniques and Types of Simulations

Cosmological simulations employ a variety of numerical techniques to accurately model the evolution of the Universe. Starting from an initial density perturbation derived from cosmological observations such as the CMB, these simulations track the distribution of matter over cosmic time [150], [151]. Specifically, "particles" are created from the matter overdensities in this initial perturbation and subsequently evolved under the influence of physical forces such as gravity. Note that these particles are not individual atoms or molecules; they are computational elements that represent a population of DM or baryonic matter, with each particle on the order of thousands to millions of solar masses. They are necessary to reduce the computational costs of cosmological simulations, which can take days or even weeks to run. The lower the mass of the particles, the higher the resolution and accuracy of the simulation.

Simulations can be grouped into two main categories based on how they treat different components of the Universe. Figure 1.6 shows the example and visualization of a few key simulations.

- **Dark matter-only (N-body) simulations**: Dark matter-only simulations, also known as N-body simulations, evolve a large number of DM particles under the influence of gravity. As the name suggests, these simulations do not include any baryonic matter such as stars, gas, and galaxies, and thus their results cannot be directly compared with observations. Despite this limitation, N-body simulations are crucial for studying the formation and evolution of large-scale structures such as clusters and the cosmic web, where baryonic processes have minimal impact. They are also computationally cheap (e.g., [152]–[154]), making them ideal for exploring a wide range of cosmological scenarios and parameters at higher resolutions.

- **Hydrodynamic simulations**: In contrast, hydrodynamic simulations, include both DM and baryonic matter. These simulations solve the equations of fluid dynamics in addition to gravitational interactions to accurately model processes like gas cooling, star formation, and supernova feedback. Hydrodynamic simulations are *significantly* more expensive than N-body simulations. This is because the hydrodynamic equations that govern baryonic processes are computationally intensive and often involve small-scale physics that require high resolution to model accurately [155]–[158]. Nevertheless, hydrodynamic simulations are critical for our understanding of galaxy formation and evolution and the interplay between DM and baryonic matter.

It is also important to further subdivide cosmological simulations based on the scales that they simulate:

- **Large-volume simulations**: Large-volume simulations capture the largest structures of the Universe such as galaxy clusters, filaments, and voids. They are thus important for testing cosmological parameters and the large-scale effects of DM and dark energy. Furthermore, consisting of millions of halos and galaxies, they are suitable statistical studies of galaxy populations. Examples include simulations like the Millennium Simulation [34] and IllustrisTNG [40], [41], which span hundreds of millions of parsecs[1] and contain billions of particles.

- **Zoom-in simulations:** As the name suggests, zoom-in simulations focus on a particular DM halo or galaxy. By increasing the resolution in a localized region of interest while maintaining a lower resolution in the surrounding area, these simulations can model the internal structure and dynamics of individual galaxies or galaxy clusters in much more detail. Examples include simulations like FIRE (Feedback In Realistic Environments, [43]), NIHAO (Numerical Investigation of a Hundred Astrophysical Objects, [157]), APOSTLE (A Project Of Simulating The Local Environment, [158]), and Auriga [156].

Limitations

Cosmological simulations are astrophysical laboratories that have allowed astrophysicists to explore the formation and evolution of the Universe in unprecedented detail. However, as with real laboratories, these simulations come with limitations. Understanding these limitations is crucial for interpreting their results accurately and for improving future simulations. Below, I discuss the key limitations of modern simulations:

- **Mass-Volume Tension**: Hinted in previous discussions, the mass-volume tension refers to the trade-off between the mass resolution and the simulated volume of cosmological simulations. High mass resolutions allow small-scale physical processes to be captured in detail, while large volumes are necessary to capture more massive (and thus much rarer) halos. However, achieving both high mass resolution and large volume simultaneously is computationally challenging and often impractical, if not impossible, even with modern technology. Large-volume simulations such as IllustrisTNG cannot accurately resolve the mass scales of dwarf galaxies [159], while zoom-in simulations such as FIRE can only simulate a small number of galaxies [160]. This limitation worsens if one wants to capture the entire formation histories of halos and galaxies up to

[1] A parsec is a common length measurement in astrophysics. A parsec is approximately 3.26 light-years.

high redshifts,[2] as structures in the earlier Universe are less massive and require higher resolution to be accurately modeled.

- **Halo Finders**: Halos and galaxies in simulations are collections of DM or star particles that are gravitationally bound together. Simulations employ the so-called "halo finders," most notably Friends-of-Friends (FoF), Subfind [161], ROCKSTAR [162], and Amiga's Halo Finder [163], to identify and characterize these structures.

  Unfortunately, halo finders have several limitations [164]–[166]. First, halos are not always spherical; they can be elongated or irregular in shape, especially when under tidal interactions with another halo. Halo finders may miss particles at the outskirts of the halo, leading to incorrect estimations of halo mass and size. Second, in dense environments, such as galaxy clusters or mergers, halo finders can struggle to distinguish between closely packed halos. Lastly, the choice of halo finder and the specific parameters used can significantly impact the resulting halo catalogs and the derived properties of halos and galaxies. The effectiveness of halo finders also depends on the resolution of the simulation; lower-resolution simulations may miss smaller halos, leading to incomplete or biased halo catalogs.

- **Baryonic Physics**: Including baryonic processes such as star formation, supernova feedback, and gas cooling introduces additional complexity and uncertainty into cosmological simulations. These processes are often modeled using subgrid physics because they occur at scales smaller than the resolution of the simulation. While these models are essential for accurately modeling galaxy formation and evolution, they are inherently approximations and *will* significantly impact the results. The accuracy of subgrid models is crucial, and discrepancies in how these processes are implemented can lead to variations in the simulation outputs. Comparisons between different subgrid physics models have been conducted in various studies to understand these impacts better. For example, Ref. [167] provide a thorough comparison of different feedback implementations and their effects on galaxy properties, while the AGORA project [168] benchmarks various hydrodynamic codes and their subgrid physics recipes.

- **Resolution and Numerical Artifacts**: The resolution of a simulation determines its ability to accurately model small-scale structures and processes. Higher resolutions require more substantial computational resources, thus limiting the cosmological volume that can be simulated, as discussed in the mass-volume tension. Low resolution

[2] Redshift refers to the phenomenon where light from distant objects is shifted to longer wavelengths due to the expansion of the Universe. Higher redshifts correspond to earlier cosmic times.

can lead to underestimation of the number of substructures and inaccurate modeling of small-scale processes. Furthermore, numerical methods used in simulations can introduce artifacts or errors. For example, the choice of time-stepping methods, grid resolution, and particle-mesh methods can affect the accuracy of the results [167], [169]. These numerical artifacts can lead to spurious features in the simulated data, complicating the interpretation of results [166], [170].

Overcoming these limitations is an important and ongoing challenge in astrophysics. For example, the mass-volume tension can be addressed by employing a multi-tier simulation approach similar to the recent GUREFT simulations, in which I participated [154]. Advances in halo finders, with recent techniques such as Symfind [166], address some of the shortcomings of earlier methods and overall improve the accuracy of halo identification and characterization. Differences in baryonic physics and numerical methods should be always treated with utmost care to ensure the reliability of simulation results. The CAMELS project [171] and the DREAMS project, to which I contributed [172], revolutionize this approach by systematically varying subgrid physics and cosmological parameters to understand their effects on simulated galaxy properties.

To conclude, cosmological simulations are indispensable tools for exploring the complex processes underlying the formation and evolution of the Universe. Despite their limitations, ongoing research efforts will continue to improve their accuracy and reliability. These improvements will advance our theoretical understanding of the complex interplay between DM and baryonic matter, bridging the gap between simulations and the real Universe.

1.2.3 Synthetic surveys

Cosmological simulations are powerful tools that can make significant theoretical predictions about our Universe. However, even without their limitations as discussed previously, they cannot be directly compared to real astrophysical observations. For example, the raw outputs of simulations are not individual stars, but particles that represent a population of thousands to millions of stars. Another example is the simulation of gas dynamics, where interactions of gas clouds and interstellar medium are represented by oversimplified models. These models may not capture all details and the different phases of the ISM, such as the cold molecular clouds, warm neutral medium, and hot ionized gas, observed in reality.

To further complicate the comparisons, observations are complex and (unfortunately) also riddled with their own systematics. Common systematics include instrumental calibration errors, atmospheric distortions, foreground contamination, and survey selection effects.

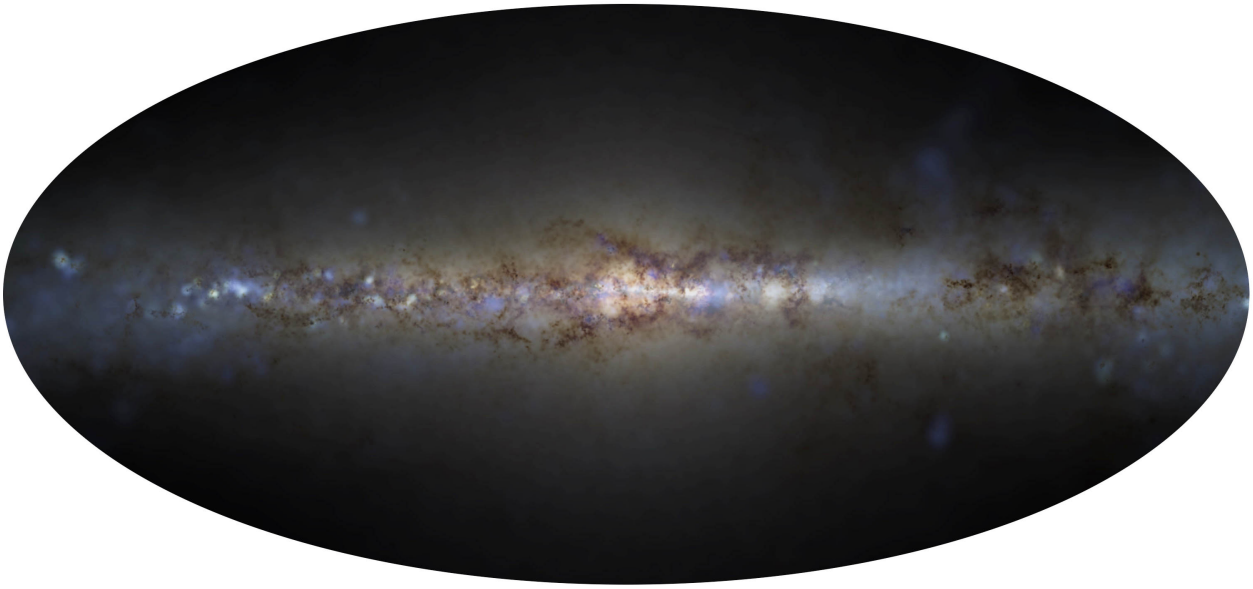

Figure 1.7: Mock real-color image of the `m12i` FIRE simulation. The synthetic survey is designed to mimic observations from the Gaia telescope and is very similar to Figure 1.3. A keen observer will notice that this galaxy is not the Milky Way because it lacks a counterpart to the Small Magellanic Cloud (SMC) and Large Magellanic Cloud (LMC). Figure adapted from [173].

These observing systematics can introduce significant biases and uncertainties in the data, further complicating comparisons with theoretical predictions.

Synthetic surveys, also called mock surveys, are simulated observations that are derived from cosmological simulations and designed to mimic real astronomical surveys (e.g., [173]–[177]). They are thus the natural next steps to bridge the gap between observations and simulations. These surveys are created by taking the output of simulations and applying observational effects to them, such as instrumental noise, resolution limits, and survey selection functions. This can involve projecting the 3D simulation data onto a 2D plane to simulate telescope images or generating stellar spectra, taking into account the characteristics of the observing instruments and even the multi-wavelength nature of real surveys, from radio waves to X-rays. Figure 1.7, adopted from Ref [174], shows an example of a synthetic survey of the Gaia telescope, generated from a Milky Way-like galaxy in the FIRE simulations.

Synthetic surveys are not without their own systematics. For example, effects such as atmospheric distortion of ground-based telescopes can be challenging to quantify and incorporate into these surveys accurately. More importantly, simulated galaxies are not perfect one-to-one matches with the Milky Way or other observed galaxies. One must assume a specific viewing point and orientation when generating synthetic observations, which can

introduce additional uncertainties. Moreover, complex treatments of dust and galactic extinction must be handled in a *self-consistent* manner to ensure realistic representations.[3] Survey effects based on the telescope's scanning law are virtually impossible to replicate. Since the slewing of the telescope depends on numerous factors, and simulated galaxies are not one-to-one matches with the Milky Way, these effects cannot be directly simulated.

Despite these limitations, synthetic surveys remain a powerful tool in astrophysics. They enable us to test theoretical models under controlled conditions, help identify potential observational systematics, and allow for the development of more robust data analysis techniques. Moreover, they serve as valuable benchmarks for upcoming observational campaigns, since they can give insights into what to expect and how to optimize the design and execution of these surveys. As observational capabilities continue to advance with new telescopes and instruments, synthetic surveys will play an increasingly vital role in interpreting the vast amounts of data collected. By iteratively improving both cosmological simulations and synthetic surveys, we can achieve a more accurate and comprehensive picture of cosmic phenomena, from the behavior of DM to the formation and evolution of galaxies.

1.2.4 Machine learning

Simulations and observations both produce incredibly complex, high-dimensional data that can be challenging to analyze and compare. For example, a stellar spectrum is high-dimensional and encodes information about the chemical compositions, temperatures, and velocities of the stars [178]. More importantly, numerous processes such as instrument noise, atmospheric interference, and foreground contamination can obscure the desired information, especially for distant objects. *As a result, astrophysical data can often have much lower signal-to-noise ratios compared to other fields of study and real-life applications.*

Moreover, the vast volume of data will only exponentially increase with future surveys. Cosmological surveys such as SDSS and DES already capture the spatial distributions of millions of galaxies [179], [180]. Meanwhile, Milky Way surveys such as Gaia capture the positions and velocities of billions of stars, along with information on their age, color, and temperature [181]–[184]. The Rubin Observatory's LSST will further revolutionize the field by observing millions of astronomical objects and producing 20 terabytes of raw data per night [113]. Many of these objects are also transients, astronomical events that change in brightness over short timescales. Their detection and analysis thus require rapid processing and immediate follow-up observations.

[3]Galactic extinction refers to the dimming and reddening of light from astronomical objects due to dust within the Milky Way. Simulated galaxies have different distributions of dust, and thus we cannot directly use the Milky Way dust map for simulations, making it challenging to model extinction accurately

Traditionally, these problems have been tackled using a variety of statistical methods. Common techniques include linear regression [185], Fourier transformation and wavelet analysis [186]–[188], and Bayesian inference (e.g., [186], [189]–[191]). Summary statistics, such as mean values and variances, are often used to condense vast amounts of data into more manageable forms. Scaling relations describe how different properties of an astronomical object correlate with each other and are used to make predictions. While these approaches have been successful, they fail to keep up with the increasing complexity and high dimensionality of modern astrophysical data. Linear regression cannot model non-linear relationships, while Fourier transform and wavelet analysis struggle to capture non-periodic or transient phenomena. Bayesian inference can be computationally intensive, and its results depend heavily on the convergence of the sampler, which can vary. In addition, to simplify the analysis process, many methods rely on Gaussian assumptions (that the data or noise follows a normal distribution) which tend to break down with more complex datasets.

The techniques described above share a common limitation: they tend to oversimplify the complexity of the data. Analyzing high-dimensional data in a reasonable computational time is seemingly an impossible task. However, when generations of astrophysics grad students look up into the night sky to seek answers, instead of solutions written in the stars, they found *machine learning*.

Overview of Machine Learning in Astrophysics

Machine learning (ML) is a branch of artificial intelligence (AI) that enables computers to learn and make decisions or predictions from data without memorization or being explicitly programmed for specific tasks. Imagine training an astrophysics graduate student to recognize stellar spectra. Instead of giving them a set of common spectra lines to identify, the advisor simply shows them many spectra of different types of stars. Over time, the graduate student learns to identify the stellar spectra based on the patterns they observe, earning them a Ph.D. Similarly, in ML, we feed an algorithm, typically a neural network, a large amount of data and train it to identify patterns and relationships within that data.

ML techniques, specifically neural networks, are designed to handle large volumes of data. They are capable of extracting subtle patterns from high-dimensional data, as has been shown in many real-life applications and other fields of science. In physics, for example, ML algorithms are employed in the Large Hadron Collider (LHC) to analyze low-level triggers and detect rare particle events [192], [193]. These capabilities make ML an invaluable asset for managing and interpreting complex datasets, and thus, it has seen increasing use in modern astrophysics. Below, I discuss some common applications:

- **Classification**: Neural networks have been used to classify astronomical objects based on their spectra, light curves, or images. For example, recurrent and Transformer-based models have been used to classify supernova light curves [194], stellar spectra [178], or gravitational waves [195]. Convolutional neural networks can be used to classify or identify desired features from in astronomical images (e.g., [196]) or gravitational-wave spectrograms (e.g. [197]). Multi-layer perceptrons have been used to classify stars that are accreted onto the Milky Way and even identify new accreted structures [65], [66]. This automated classification significantly speeds up the analysis process and allows us to handle the vast amount of data generated by modern surveys.

- **Regression and Inference**: Similarly, neural networks can be used to predict continuous trends based on the input data (regression) or draw conclusions about the underlying parameters that govern these trends (inference). In astrophysics, they have been used to predict properties such as the redshift, stellar mass, and star-formation rates of galaxies from multi-band imaging data [198]–[200]. For inference, ML models have been employed for various applications ranging from extracting cosmological parameters from large-scale structures [201]–[204], DM properties from stellar motions and strong lensing images [93], [205], [206], to properties of gravitational-wave progenitors [207].

- **Clustering and Anomaly detection**: Neural network architectures such as variational autoencoders (VAE) are commonly employed for clustering and anomaly detection. These architectures learn to group similar data samples, typically by minimizing the distance in their representation (latent) space during training. Most deep learning algorithms for clustering also give access to this latent space, allowing them to be used to detect anomalous samples. For example, ML have been used to identify unmodeled transient gravitational waves (e.g., typically non-merger sources such as supernovae) [208] and anomalous images in optical surveys [209]. Clustering techniques are used to identify structures in data or data compression (e.g., [210], [211]).

- **Generative models (emulators)**: Generative models, also called emulators, are designed to learn the properties of observed or simulated data and generate new samples. They can be used to fill in gaps in parameter spaces, predict future observations, or even create entirely new datasets for training and testing other ML algorithms. For example, Generative Adversarial Networks (GANs) have been used to generate density fields in cosmological simulations and "paint" galaxies onto DM-only fields [212], [213]. VAEs are commonly used to "de-noise" noisy observations, by learning to reconstruct clean samples from noisy samples (e.g., [214], [215]) or reconstruct missing information

(e.g., [216]). Normalizing flows [217] have been used to predict the galaxy mass functions and other relations vary with cosmological parameters [218]. Diffusion models have been used for point-cloud modeling of cosmological simulations [219].

Limitations

While ML techniques have shown great promise in astrophysics, they are not without limitations. In this section, I list a few key limitations:

- **Data Quality and Quantity**: High-quality, labeled data is critical for training robust ML models. Acquiring such data, however, can be challenging, especially in astrophysics due to observational limitations, noise, and incomplete datasets. For example, astronomical phenomena such as supernovae are extremely rare and difficult to build a training dataset for.

- **Simulations versus Reality (Generalization)**: To tackle the data challenge, astrophysicists often rely on simulations, whether they are cosmological simulations that model entire universes, template-based simulations of transients, or even numerical relativity simulations of gravitational waves. However, simulated data can never perfectly match real observations due to simplifications and assumptions inherent in the models. Additionally, as also discussed previously, observational systematics and uncertainties are difficult, if not impossible, to model. Discrepancies between simulations and actual data can lead to subtle biases in the trained models. It is thus extremely important to continue improving simulation techniques to better reflect the complexities of the Universe. Domain adaptation techniques [220]–[222], which address shifts in data distribution, are also a significant focus of research in the ML community.

- **Uncertainty quantification**: The challenge of reconciling simulations and reality is not unique to ML techniques. For example, statistical methods that are based on template matching will share similar limitations if the templates do not perfectly match real data. However, this issue is exacerbated for ML models because these models often struggle to provide clear measures of uncertainty, making it challenging to assess the reliability of their predictions.

  It is worth dividing uncertainties into two categories: *aleatoric uncertainty*, which is the inherent noise in the data, and *epistemic uncertainty*, which reflects the uncertainty due to the model's lack of knowledge or incomplete understanding of the underlying data distribution. Aleatoric uncertainty can be addressed using techniques like normalizing flows and simulation-based inference, which can model the inherent noise in

the data. Epistemic uncertainty can be tackled with methods such as Bayesian neural networks and ensemble learning, which provide a measure of the model's confidence in its predictions and can highlight areas where the model lacks sufficient information. Properly addressing these uncertainties is crucial to ensuring the robustness and reliability of ML applications in astrophysics.

- **Interpretability**: ML models often operate as "black boxes," providing predictions without clear explanations of how those predictions were made. This lack of interpretability can be problematic in astrophysics research, where understanding the underlying processes and validating results are important. Additionally, this exacerbates the generalization problem discussed above. However, it is worth noting that it is not impossible to gain insights into these models. They can often be subjected to stress tests to evaluate their reliability and robustness. For many applications, it is practical to use these models even without knowing the exact mechanisms behind their predictions, as long as their outputs are validated and their limitations are well-understood. The reliability and interpretability of neural networks are among the most active areas of research in the ML community [223]–[225].

- **Overfitting**: Overfitting occurs when a model learns to capture features or noise uniquely present in the training data rather than the underlying signal. This can lead to poor generalization to new, unseen data. ML models are more susceptible to overfitting because they are too complex relative to the amount of training. However, in many cases, this can be easily identified and addressed with more training data, better regularization techniques (e.g., dropout, batch normalization, early stopping, learning rate scheduler [226]–[228]), or simply reducing the complexity of the model. It is especially important to note that many ML models follow a well-documented *double descent phenomenon*, which describes a counter-intuitive behavior where increasing model complexity initially leads to worse performance before eventually improving it again as complexity continues to rise. This also partially explains why ML techniques tend to have better performance than traditional methods that have fewer parameters. Interested readers should refer to Refs. [229], [230].

- **Computational resources**: Deep learning models require significant computational power and memory to train. Additionally, the environmental impact of ML models is an emerging area of concern, primarily due to the substantial energy consumption and carbon emissions associated with training and deploying these models (see e.g., https://aiindex.stanford.edu/report/). Thus, one should also take special care to optimize and manage resources effectively to ensure efficient model training and deployment.

Addressing these limitations requires a combination of improved data acquisition techniques, the development of more interpretable and robust model architectures, the integration of domain knowledge, and advanced methods for uncertainty quantification. *Ongoing research and collaboration between astrophysicists, ML scientists, and industry professionals are crucial for overcoming these challenges and realizing the full potential of ML in astrophysics.*

The Next Steps

Despite these limitations, ML techniques remain a powerful tool in advancing astrophysical research. In this section, I highlight a few exciting areas of ML research, in no particular order:

- **Simulation-based Inference**: Simulation-based inference determines the parameters of a model by comparing observed data with data generated through simulations (see Ref.[231] for a review). Unlike traditional techniques, simulation-based inference can directly leverage the simulations without the need for an explicit likelihood.[4] Key techniques include conditional neural density estimation methods such as conditional normalizing flows [217], [232]. As also discussed previously, simulation-based inference is powerful for scientific applications as it can model the aleatoric uncertainty, or the uncertainty inherent in the simulations and data. Simulation-based inference has found many applications in astrophysics, including strong lensing [205], [206], cosmology [202], [203], and DM studies [93], [201].

- **Diffusion models**: Diffusion models are the next generation of generative modeling. These models work by gradually corrupting the data with noise and then learning to reverse this process to recover the original data distribution, [233]–[236]. They have many inherent advantages over other generative models, such as GANs, VAEs, and normalizing flows. Unlike GANs, diffusion models allow both sampling and likelihood evaluation, making them much more suitable for scientific applications. Also unlike GANs, they are much more stable to train and also less susceptible to mode collapse [237], [238]. Diffusion models are also more expressive and allow for more flexible neural network architectures than VAEs and normalizing flows. For example, diffusion models have been employed mainly for generating galaxy surveys from large-volume simulations [219], density fields [239], [240], and radio images [241].

[4]For this reason, it is also called likelihood-free inference, though this is a bit of a misnomer because the likelihood is built in implicitly through the simulations.

- **Geometric deep learning**: Many astrophysical observations inherently reside in non-Euclidean spaces. For example, the coordinates of celestial objects are often given in Right Ascension and Declination, which are analogous to longitude and latitude on a sphere. Leveraging this spherical geometry is key to accurately modeling and analyzing the spatial distribution of celestial objects. Geometric deep learning employs principles of geometry and topology to handle data in these non-Euclidean spaces. Techniques such as graph neural networks [242], [243] can effectively represent and process data on computational graphs. For example, graph neural networks have been used in conjunction with simulation-based inference to infer cosmological and dark matter parameters from the distribution of galaxies or star clusters (e.g., [93], [202]–[204]).

  Additionally, geometric deep learning techniques often make it easier to incorporate geometric invariants such as rotation, translation, or even Lorentz invariance and equivariance [244][5]. This is particularly powerful for physics and astronomy applications, where symmetry plays a crucial role in many physical processes.

- **Multi-modal models**: Astrophysical observations come in many forms and flavors. A galaxy, for example, can be observed through its images across different wavelengths, its full spectra, and its light curves showing its brightness variations over time. Each of these data types (known as "modes" to the ML community) encodes different but somewhat overlapping information about the same galaxy. Multi-modal models can exploit information from different modalities, often by using self-supervised learning to map these diverse data types into the same latent space. This unified latent representation allows these models to leverage the strengths of each mode and often leads to better generalizations.

  It is worth emphasizing that these models are designed to still work effectively even when only one data type is accessible. In many of these cases, their performance has been found to even be superior to non-multi-modal models, though the exact reasons for this improved performance are still an active area of research in the ML community [245]–[248]. An exciting recent work introduced AstroCLIP [199], a model that creates a shared embedding space for galaxy images and spectra, and demonstrated better performance on tasks like redshift estimation and galaxy property prediction, even compared to supervised baselines. Note that each data mode does not have to cor-

[5]Invariance refers to a property of a model where the output remains unchanged when the input undergoes certain transformations. Equivariance means that the output of the model transforms in a predictable way when the input is transformed. For example, a rotation-invariant model would produce the same output regardless of the rotation of the input, while a rotation-equivariant model's output would rotate in the same manner as the input.

respond to different data types; multi-modal methods can also be applied to the same data type of different telescopes and instruments, such as the same images taken by two different telescopes. This opens up for more robust applications of ML techniques on astronomical data.

- **Foundational models**: Foundational models are large-scale, pre-trained models designed to be fine-tuned for various downstream tasks. They provide a robust starting point by leveraging vast amounts of data and powerful architectures, making them adaptable to a wide range of applications in astrophysics, from image recognition to spectral analysis. Examples of recent foundational models in astrophysics include AstroCLIP [199] and Radio Galaxy Zoo [249].

1.3 Roadmap and the Big Picture

1.3.1 Roadmap

In the previous section, I discuss how ML techniques are powerful analysis tools that have increasingly seen more applications in astrophysics and cosmology. While they have the potential to transform astrophysics research, they come with their own set of limitations and challenges. Addressing these challenges is crucial for making significant astrophysical discoveries and is an active area of research. In this thesis, I present a few exciting applications that showcase the power of ML in DM research, galaxy formation, and cosmology. My thesis is structured as follows:

- In Chapter 2, I present a novel application of graph neural networks and simulation-based inference to constrain the DM density profiles of dwarf galaxies from the kinematics of their stars. This approach provides stronger constraints on the density profiles than the traditional Jeans dynamical modeling techniques, thereby addressing the core-cusp discrepancy and other small-scale structure issues in DM halos. This work was published in *Physics Review D* Volume 107, Issue 4, Page 043015 in February 2023 [93]. Though the method presented in this chapter has only been mostly applied to simple, analytical simulations, in my postdoctoral research, I will extend it to more sophisticated, cosmological simulations and eventually real data.

- Chapter 3 presents FLORAH, a generative model for mass assembly histories of halos, based on recurrent neural networks and normalizing flows. The mass assembly histories are crucial for understanding galaxy formation and evolution, yet current analytic methods fall short in accuracy and fail to capture their relationship with halo

structure and environment. I demonstrate FLORAH's ability to accurately model the mass assembly histories, as well as other key relations such as the stellar-to-halo mass relations, concentration-mass relations, and even galaxy assembly bias. This work was submitted to *Monthly Notices of the Royal Astronomical Society* and is currently under peer-reviewed. A copy is available on arXiv at https://arxiv.org/abs/2308.05145 [250]. Future work will focus on extending FLORAH to emulate the entire merger histories, which can then be used for small-scale applications such as generating mock tidal disruption structures.

- Chapter 4 combines normalizing flows, variational diffusion models, and Transformer-based neural networks to emulate the satellite populations under WDM scenarios. By modeling satellite populations as point clouds, this method, called NeHOD (Neural Halo Occupancy Distribution), can uniquely emulate the kinematics and spatial distributions of these satellites, which other methods have struggled with. I train NeHOD on the WDM Milky Way suites of the DREAMS simulations and show that it can capture how key summary statistics, such as satellite counts, mass functions, concentration-mass relations, etc. vary with the WDM particle mass and astrophysical parameters. The work is currently under internal revision of the DREAMS collaboration before being submitted to *The Astrophysical Journal*.

- In Chapter 5, I present the *Ananke* synthetic surveys of 3 Milky Way-like galaxies in the FIRE simulations. These surveys are designed to match observations from the Gaia DR3 (Data Release 3) catalog [181]–[184], which was released on June 13, 2022. Prior work with the Gaia DR2 catalog was led by Robyn Sanderson in Ref. [173]. In this work, my collaborators and I extend *Ananke* to Gaia DR3 by incorporating observational effects such as measurement uncertainties, dust extinction, and survey selection effects. I share an equal contribution with Xiaowei Ou, a fellow Ph.D. student in Lina's group. This work was published in *The Astrophysical Journal* Volume 966, Number 1, Page 108 in April 2024 [174].

- Finally, I conclude my thesis in Chapter 6.

1.3.2 The Big Picture

In this section, I briefly discuss how the above projects tie into the bigger picture of understanding the properties of DM and the physical process behind galaxy formation and evolution.

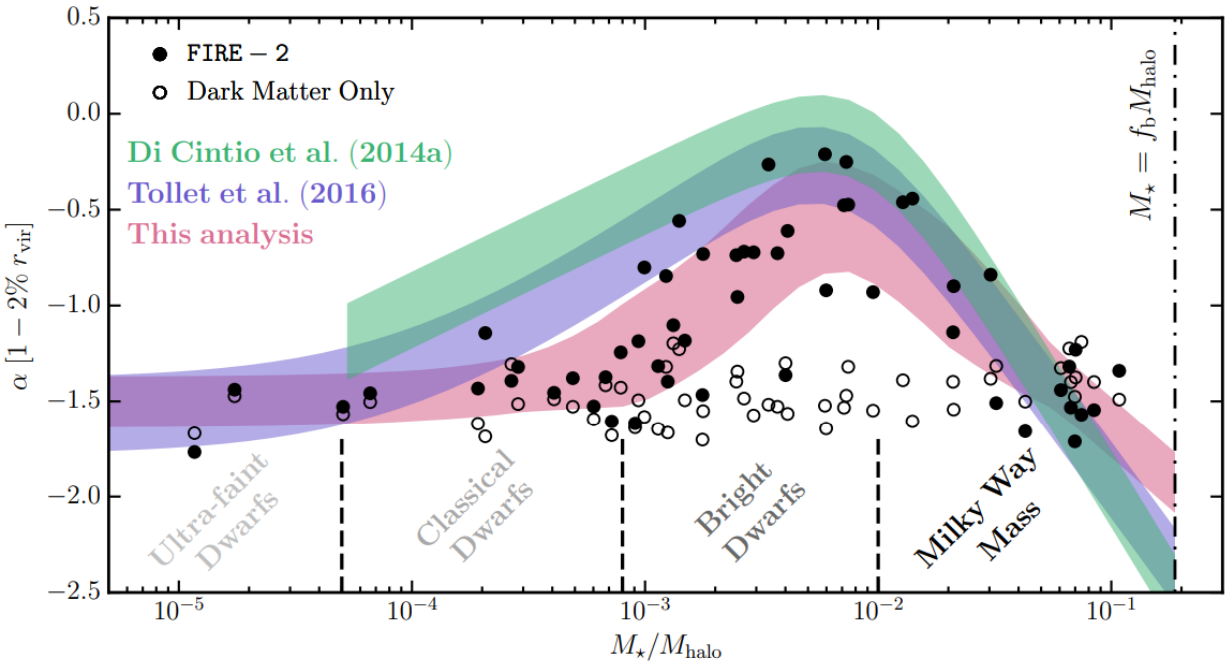

Figure 1.8: The inner slope α of the DM density profiles as a function of the stellar-to-mass ratios of galaxies in the FIRE-2 simulations. Here, $\alpha = 0$ denotes cored profiles, while $\alpha = -1$ denote cuspy profiles. We see the impact of stellar feedback, characterized by the mass ratios, on the density profiles. Figure adapted from [251].

The work described in Chapter 2 and Chapter 4 will help place constrains on both DM properties such as the annihilation cross-section of DM and the strength of baryon feedback. The work described in Chapter 3 will help create more accurate theoretical models for galaxy formation and evolution, which can then be compared to results from JWST. The work described in Chapter 5 can be used to create mock catalogs and training datasets for current astrophysical surveys, as well as forecasts for future ones, thus bridging the gap between observations and theoretical modeling.

Figure 1.8, adapted from Ref. [251] shows how the inner slope α of the DM density profiles vary as a function of the strength of the feedback, which is characterized by the fraction of stellar mass $M_\star/M_{\rm halo}$, of galaxies in the FIRE-2 simulations [43], [253]. Measurements of the inner slope α in real dwarf galaxies ($M_\star/M_{\rm halo} < 10^{-3}$) rely on the stellar velocity dispersions and Jeans equations and are relatively under-constrained [117]–[119]. In addition, Ref. [254] shows that these techniques cannot distinguish between cored ($\alpha = 0$) and cuspy ($\alpha = -1$) density profiles with 1,000 stars (see Table 3 and 4 in Ref. [254]). This creates challenges in comparing the values of α from observations with the theoretical fits from hydrodynamic simulations in Figure 1.8. With the simulation-based inference framework presented in Chapter 2, I will place a better constrain on the inner slope α for dwarf galaxies,

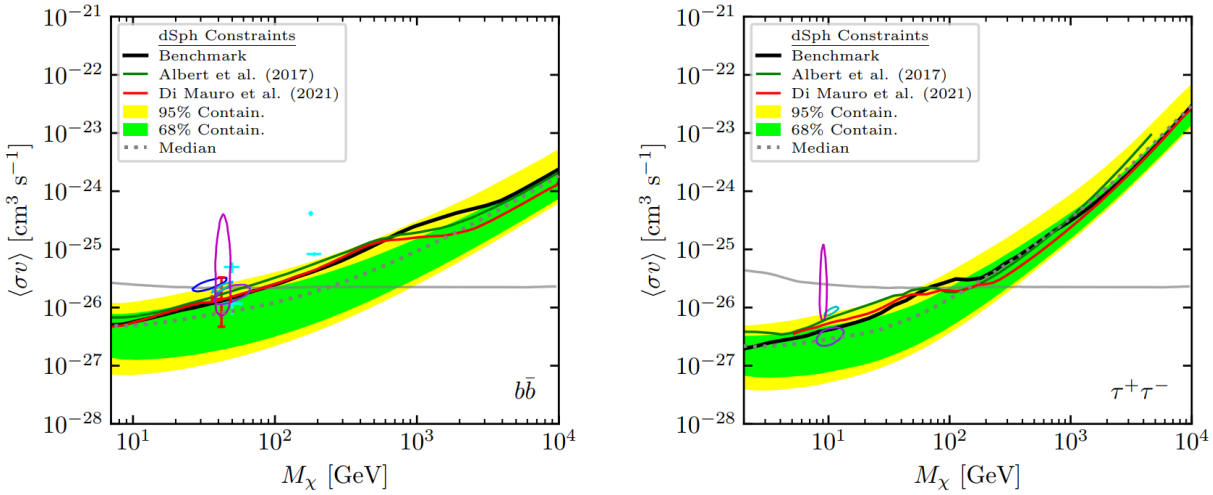

Figure 1.9: Constraints on the $b\bar{b}$ (left) and $\tau^+\tau^-$ (right) annihilation channels derived from Fermi-LAT data from Ref. [252]. Figure adapted from [252].

especially ultra-faint dwarfs with few stars, which will facilitate the comparisons between observations and simulations. In addition, constraining the DM density profiles will also improve estimations of the so-called J-factor, which is proportional to the expected signal flux from DM annihilation:

$$J = \int \mathrm{d}s \int \mathrm{d}\Omega\, \rho^2(s, \Omega). \tag{1.3}$$

where ρ is the DM density profiles, s is the line-of-sight distance, and Ω is the solid angle. Accurately determining J-factors are important for DM indirect detection experiments that use dwarf galaxies as targets [88], [94]–[101]. Figure 1.9, adapted from Ref. [252], shows the upper limits derived from Fermi-LAT data on the annihilation cross-section of DM for the $b\bar{b}$ and $\tau^+\tau^-$ channels, where b and τ denote the bottom quark and tau particle, respectively. The containment regions, shown as the yellow and green bands, are calculated from Monte Carlo simulations assuming a certain background and the measured values of the J-factor. The black line shows the upper limit derived from observations of dwarf galaxies. If the observed upper limit is greater than the containment regions, this suggests a potential excess of events that could be indicative of DM interactions at that particular mass and cross-section. The widths of these containment regions are influenced by the uncertainties in the J-factor. Therefore, improving the precision of J-factor measurements would enhance the sensitivity of these searches. The simulation-based framework in Chapter 2 aims to achieve that by providing more accurate constraints on the DM density profiles and subsequently the J-factor. For example, in the preliminary analysis on idealized data, we found that our framework can, on average, reduce the error bar on the J-factor by approximately 10-20%.

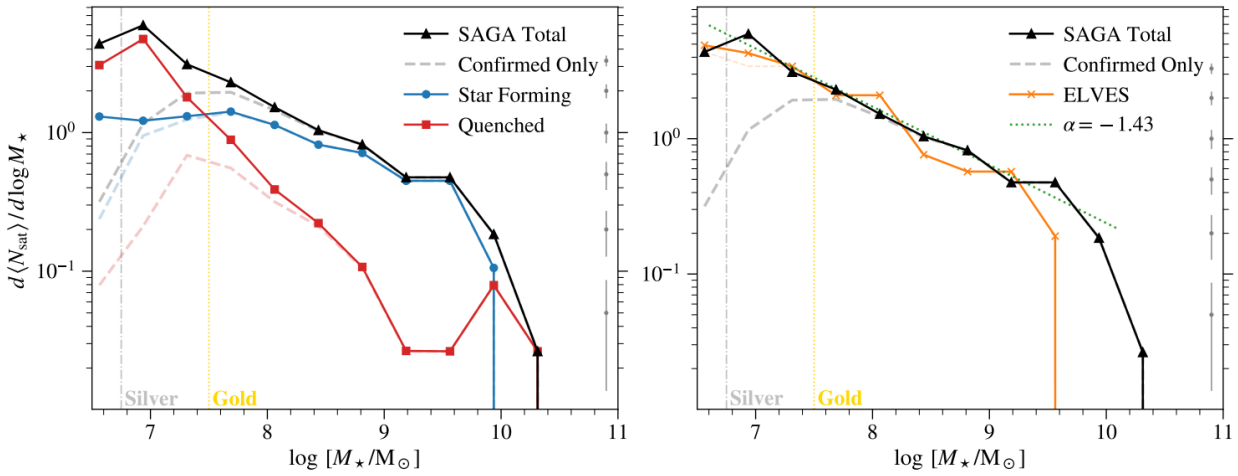

Figure 1.10: The satellite stellar mass functions in differential form reported by the SAGA survey [80]–[82]. The left panel shows the mass functions for two populations of satellite galaxies, separated by their star formation rate. The right panel compares the SAGA mass functions with previous results from ELVES [105]. Figure adapted from [80].

Chapter 4 presents NeHOD, a generative model for satellite galaxies in Milky Way-like halos. NeHOD can be thought of as a machine learning-based extension to Halo Occupation Distribution (HOD) methods, in which galaxies are assigned to DM halos in N-body simulations based on a statistical framework that describes the probability of a halo hosting a certain number of galaxies [255]. This process, known as "painting" galaxies onto DM halos, follows a set of recipes that aim to capture the main physics describing the halo-galaxy connection. HODs have been applied in various contexts, from studying galaxy clustering [256]–[258] and the halo-galaxy connection [259]–[261] to generating mock galaxy catalogs for recent surveys such as DESI [262] and Euclid [263]. NeHOD can be used to predict properties such as satellite mass functions, which can then be compared with observations from dwarf galaxy surveys such as SAGA [80]–[82] and ELVES [105]. Figure 1.10, adapted from Ref. [80], shows how star formation affects the satellite stellar mass functions in the SAGA survey. Note that the satellite stellar mass functions will also be impacted by the DM models. For example, WDM models predict fewer low-mass subhalos due to free-streaming in the early Universe and thus also affect the satellite mass functions in a similar way. Therefore, an accurate theoretical modeling of satellite mass functions, combined with other independent observations like the ones described above, is the key to disentangle and place constraints on both the DM properties and baryonic processes. The NeHOD framework presented in Chapter 4 is a promising step towards this.

Chapter 3 presents FLORAH, a generative model for assembly histories of DM halos. Assembly and merger histories of halos are often combined with semi-analytic models (SAMs)

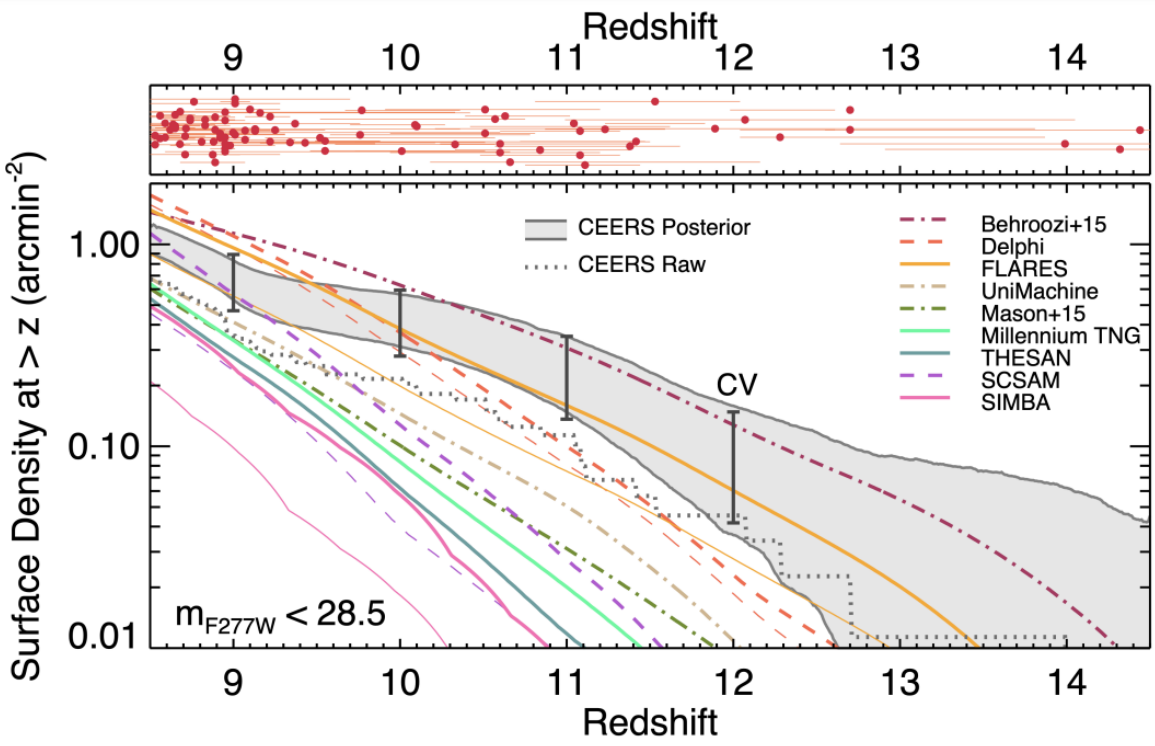

Figure 1.11: The top panel shows the observed samples of galaxies as a function of redshift by the CEERS [111], [264] survey of JWST. The bottom panel shows the cumulative surface density of bright galaxies as a function of redshifts. The F277W denotes the wide-band mid-infrared (MIR) filter with the central wavelength of approximately 2.77μm. The gray band shows observations of the JWST. The lines show the theoretical predictions from cosmological simulations (solid lines), semi-analytical models (dashed lines) and semi-empirical models (dashed dotted lines). We see that theoretical models under-predict the abundance of high-redshift galaxies, which is characterized by the surface density. Figure adapted from [264].

to predict the formation and evolution of galaxies [265]–[287]. For example, SAMs have been used to make predictions on the abundance of high-redshift galaxies, which are important for interpreting results from JWST and placing constraints on the formation processes of the first galaxies. Figure 1.11, adapted from Ref. [264], shows an example result from the CEERS (Cosmic Evolution Early Release Science) survey of JWST. In the Figure, we see a disagreement in the number of high-redshift galaxies observed by JWST (gray bands) and predicted by theoretical models, including cosmological simulations, semi-analytical models, and semi-empirical models. Specifically, theoretical models under-predict the number of high-redshift galaxies. This creates a tension between observations and theories of galaxy formations and potentially even cosmology. A potential way to alleviate this tension is to have a more accurate model of the formation of supermassive black holes in the early Universe. The FLORAH framework presented in Chapter 3 will generate fast and accurate mass assembly histories of halos, which are an important ingredient for SAMs. This will facilitate testing of theoretical models for supermassive black holes seeding, thus improving the accuracy of SAMs.

Chapter 2

Uncovering dark matter density profiles in dwarf galaxies

Disclaimer

The content of this chapter was published in *Physics Review D* Volume 107, Issue 4, Page 043015 in February 2023 [93]. The authors are Tri Nguyen, Siddharth Mishra-Sharma, Reuel Williams, and Lina Necib.

Abstract

Dwarf galaxies are small, dark matter-dominated galaxies, some of which are embedded within the Milky Way. Their lack of baryonic matter (e.g., stars and gas) makes them perfect test beds for probing the properties of dark matter—understanding the spatial dark matter distribution in these systems can be used to constrain microphysical dark matter interactions that influence the formation and evolution of structures in our Universe. We introduce a new method that leverages simulation-based inference and graph-based machine learning in order to infer the dark matter density profiles of dwarf galaxies from observable kinematics of stars gravitationally bound to these systems. Our approach aims to address some of the limitations of established methods based on dynamical Jeans modeling. We show that this novel method can place stronger constraints on dark matter profiles and, consequently, has the potential to weigh in on some of the ongoing puzzles associated with the small-scale structure of dark matter halos, such as the core-cusp discrepancy.

2.1 Introduction

Cosmological structure formation is known to proceed hierarchically—smaller structures seed the formation of larger structures [288]. Dark matter (DM) plays an outsized role in this process, acting as a "scaffolding" on which structure evolution plays out. At the same time, the precise mechanism of structure formation is keenly sensitive to the microphysical properties of DM e.g., the nature of its self-interactions. Deviations from the canonical Λ Cold Dark Matter (ΛCDM) paradigm of cosmology would be imprinted in the properties of DM clumps (known as *halos*) on smaller spatial scales. Robustly characterizing the distribution of small-scale structures in our Universe may therefore hold the key to answering one of the major unsolved questions in particle physics and cosmology—the particle nature of DM.

Dwarf galaxies are small galaxies, some of which are embedded within larger galaxies like the Milky Way. They are dominated by DM [15], making them versatile astrophysical laboratories for DM studies. A major goal in cosmology and particle physics is to detect non-gravitational interactions of DM. For the canonical Weakly-Interacting Massive Particle (WIMP) DM paradigm, one of the main avenues to do so is DM indirect detection: WIMPs could annihilate into Standard Model (SM) particles, producing striking signatures from DM-overdense regions in γ-ray observations [88], [94]–[101] Being deficient in baryonic matter, dwarf galaxies act as ideal targets for indirect detection, with a relatively large predicted ratio of DM signal to astrophysical background.

A pervasive puzzle in cosmology is the so-called core-cusp discrepancy, referring to the question of whether the inner DM density profiles of dwarf galaxies are cuspy (steeply rising) or cored (flattened) [289], [290]. N-body simulations using ΛCDM cosmology suggest that in the absence of baryonic physics, cold DM halos follow the cuspy Navarro-Frenk-White density (NFW) profile [117], which is characterized by a steep rise in the density $\rho \propto r^{-1}$ at small halo-centric radii r. However, recent measurements of stellar dynamics suggest that these systems could instead have a flattened density profile at their center, also known as a *core* [117], [118]; see Ref. [45] for a review. Potential solutions to the core-cusp discrepancy range from stellar feedback which ejects baryons and flattens the DM central density profile [72]–[75] to alternative DM models like self-interactions [76]–[78].

DM density profiles in dwarf galaxies are traditionally inferred using spectroscopic observations of line-of-sight velocities and angular positions of stars gravitationally bound to these systems. In particular, integral moments of the Jeans equation can be used to relate the velocity dispersions of tracer stars to the gravitational potential of the system [91], [291]. Although Jeans modeling has proven highly successful for modeling DM distributions in dwarf galaxies, there are several caveats and limitations associated with this approach (see

e.g. Refs. [85], [254], [292]). For example, Jeans modeling assumes the system is in dynamical equilibrium, which may not be a robust assumption given the active merger history of the Milky Way (see Ref. [106] for a review). Assumptions such as isotropy of the gravitating system are also often necessary in order to enable a tractable analysis. Finally, by relying on a simplified description of the data through second moments of the stellar velocity distribution, inference based on Jeans modeling is likely to lose some of the salient information available in observations. Absent additional assumptions, this is expressed as a degeneracy between the mass profile of the system and the anisotropy structure of stellar orbits, known as the mass-anisotropy degeneracy. Exploiting additional information about the modeled stellar phase-space distribution can further inform the DM density profile, helping break this degeneracy. Several methods have been proposed in the literature to this end [293], including using higher-order moments of the Jeans equation [92], [292], [294], leveraging multiple distinct tracer populations [295]–[297], including measured proper motions when these are available [298], [299], and alternative strategies to solving the Jeans equation [300].

This chapter introduces a new machine learning-based approach for linking observed stellar properties to the DM density profiles of dwarf galaxies. Our method is based on forward modeling simulated dwarf galaxy systems and corresponding observations, learning to summarize representative features from these data sets using graph neural networks, and performing simulation-based inference to simultaneously extract the spatial profiles associated with the DM and stellar components of the dwarf galaxy.

2.2 Methodology

We describe, in turn, the forward model used in this study and its realization via simulations, the representation of stellar kinematic data as a graph, and finally the feature-extractor graph neural network and simulation-based inference procedure. We show a schematic illustration of our method, including a rough breakdown of the different steps of the pipeline, in Figure 2.1.

2.2.1 Datasets and the forward model

In this proof-of-principle exposition, we consider idealized simulations of spherical and dynamically equilibrated dwarf galaxies. Our forward model is fully specified by the joint distribution function (DF) $f(\vec{x}, \vec{v})$ of positions and velocities of stars following a certain (a-priori unknown) spatial distribution (known as the *light profile*). These tracer stars are gravitationally bound to a DM halo with a density profile we wish to infer. We use the public

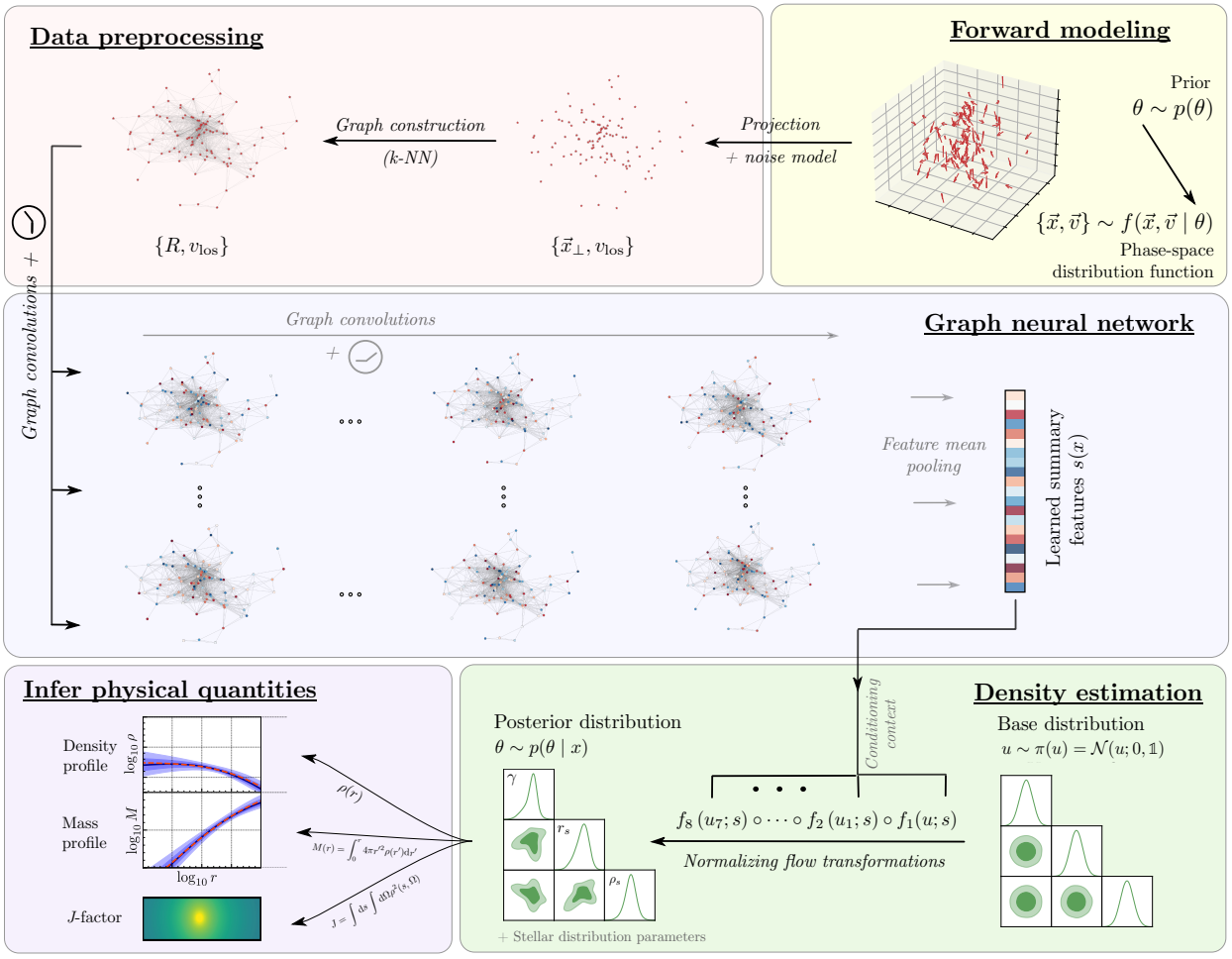

Figure 2.1: A schematic illustration of our method for inferring DM density profiles of dwarf galaxies from observed stellar kinematics.

code `StarSampler`[1] to generate simulated realizations of stellar kinematics (6-D position and velocity phase-space components) from the forward model. `StarSampler` uses importance sampling [301]–[303] to sample the DF of positions and velocities of tracer stars in a given DM potential.

We model the DM profile using the generalized Navarro–Frenk–White (gNFW) profile [117]:

$$\rho_{\rm DM}^{\rm gNFW}(r) = \rho_0 \left(\frac{r}{r_s}\right)^{-\gamma} \left(1 + \frac{r}{r_s}\right)^{-(3-\gamma)}, \tag{2.1}$$

which depends on three free parameters: the density normalization ρ_0, the scale radius r_s, and the asymptotic inner slope γ. The inner slope value $\gamma = 1$ corresponds to a cuspy NFW profile, while $\gamma = 0$ corresponds to a pure DM core. We consider these two parameter points as benchmarks, since the ability to robustly distinguish between the two possibilities would offer a path towards resolution of the core-cusp discrepancy.

We assume a stellar density distribution $\nu(r)$ that follows a 3-D Plummer profile [304]:

$$\nu(r) = \frac{3L}{4\pi r_\star^3} \left(1 + \frac{r^2}{r_\star^2}\right)^{-5/2} \tag{2.2}$$

where L is the total luminosity and $r_\star$ is the scale length. We also introduce a velocity anisotropy profile $\beta(r)$ in order to model deviations from circular orbits; $\beta(r)$ is defined similarly to Ref. [305], [306] as

$$\beta(r) = \frac{r^2}{r^2 + r_a^2}, \tag{2.3}$$

which has an additional parameter r_a describing the radius of transition from isotropic velocity orbits at small radii to radially-biased orbits at larger radii.

In total, our model has three DM parameters (ρ_0, r_s, γ) and two stellar parameters $(r_\star, r_a)$. Assuming the gravitational potential of the system is dominated by DM, and the model is independent of L in Equation 2.2. We provide further details of the forward model, phase-space distribution function, and a summary of the prior distributions of the DM and stellar parameters in Appendix A.1.

We generate 80,000 training samples, 10,000 validation samples, and 10,000 test samples using the prior parameter distributions. Each sample contains the 3-D positions and 3-D velocities of tracer stars with respect to the center of a dwarf galaxy. The number of stars in each galaxy is sampled from a Poisson distribution $n_{\rm stars} \sim {\rm Pois}(\mu_{\rm stars})$. We set $\mu_{\rm stars} = 100$ stars in our baseline benchmark, roughly corresponding in order of magnitude to the number of stars typically observed in dwarf galaxies of interest [18], [89], [90], [307].

[1] https://github.com/maoshenl/StarSampler

We explore variations on this choice in Appendix A.2.

2.2.2 Data pre-processing and graph construction

For each kinematic sample, we randomly draw a line-of-sight axis and project the galaxy onto the 2-D plane perpendicular to it. We then derive the 2-D projected spatial coordinates with respect to the center of the galaxy (X, Y) and line-of-sight velocities $v_{\rm los}$ for each star in these coordinates, reflecting typically-available observations. To study the validity of the method before the inclusion of large measurement errors, we assume a Gaussian velocity noise model with standard deviation 0.1 km/s. The effect of larger measurement errors is shown in Appendix A.2.

We can represent the stellar kinematic data in the form of a potentially weighted, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V}$ is a set of nodes representing $|\mathcal{V}| = N_{\rm stars}$ individual stars, $\mathcal{E}$ is a set of edges, and $A \in \mathbb{R}^{N_{\rm stars} \times N_{\rm stars}}$ is an adjacency matrix describing the weights of connections between vertices. This representation is well-suited for our purposes since the stars in a dwarf galaxy have no intrinsic ordering, and the graph structure can efficiently capture information about the phase-space correlation structure, including higher-order moments [204].

In our analysis, each node represents a star, with the node features being its line-of-sight velocity $\tilde{v}_{\rm los}$ and the projected radius $R = \sqrt{X^2 + Y^2}$. We choose to use R instead of the full (X, Y) coordinates in order to incorporate projective rotational invariance into the graph representation, which was found to enhance the simulation-efficiency of our method.

To determine the graph edges $\mathcal{E}$, we calculate pair-wise distances between all stars using (X, Y), then connect each star to its k-nearest stars including itself. Since the edges are assumed to be undirected, each star can be connected to more than k other stars. We found $k = 20$ to provide a good trade-off between model performance and computational overhead. Finally, we do not include edge weights in our graph, however, we have experimented with a variety of weighting schemes, including attention-based learned weights [242], and found them to perform similarly in downstream inference to the unweighted case.

2.2.3 Neural network architecture and optimization

We use a graph neural network (GNN) $g_\varphi : \mathcal{G} \to \mathbb{R}^{N_{\rm feat}}$ in order to extract $N_{\rm feat}$ summary features from the constructed graph representation $x \in \mathcal{G}$ of mock dwarf galaxy stellar kinematic data. Here φ represent the parameters of the graph neural network. The feature-extraction network consists of 5 graph-convolutional layers, each with 128 channels, based on convolutions in the Fourier domain using a basis of Chebyshev polynomials of order

4 as filters [308]. This is followed by a global mean pooling layer which aggregates the permutation-equivariant features into a permutation-invariant representation, and a fully-connected layer which projects the output onto a set of $N_{\rm feat} = 128$ summaries $g_\varphi(x)$. We explore variations on the graph-convolution scheme in Appendix A.2.

The joint posterior $\hat{p}_\phi(\theta \mid g_\varphi(x))$ of the five parameters of interest θ characterizing the DM and stellar profiles is modeled using a normalizing flow [232], [309]—a class of flexible generative models that allow for efficient density estimation as well as sampling. The flow transformation (with parameters ϕ) is conditioned on the summary features extracted by the graph neural network and its negative log-density $-\log \hat{p}_\phi\left(\theta \mid g_\varphi(x)\right)$ is used as the optimization loss. Our flow model consists of 4 Masked Autoregressive Flow (MAF) transformations, each using a 2-layer Masked Autoencoder for Distribution Estimation (MADE) with hidden dimension 128 [310]. This method falls under the class of approaches known as simulation-based inference (see Ref. [311] for a review), specifically neural conditional posterior estimation [312], [313].

The GNN and normalizing flow parameters $\{\varphi, \phi\}$ are optimized simultaneously on the 80,000 simulated training samples using the AdamW optimizer [314], [315] with an initial learning rate of 5×10^{-4} and a weight decay of 10^{-2} using a batch size of 64. At the end of each epoch, we evaluate the loss on the 10,000 held out validation samples and reduce the learning rate by a factor of 10 if no improvement is seen after 4 epochs. We stop training if the validation loss has not improved after 10 epochs. Model training typically terminates after $\sim\ 30 - 40$ epochs, which takes ~ 30 minutes on a single NVIDIA Tesla V100 GPU.

2.3 Results and Conclusion

We apply our pipeline to 10,000 test dwarf galaxies and summarize our results in Figure 2.2. For each galaxy, we condition the trained normalizing flow on features extracted using the trained GNN feature extractor and draw 10,000 samples from the joint DM and stellar posterior. We then compute the marginal medians as the predicted parameters and sort them into bins based on their truth values. Figure 2.2 shows the median (solid blue line), middle-68% (blue bands), and middle-95% (light blue bands) credible intervals for each bin of the DM parameters. In general, our method successfully recovers individual DM parameters consistent with the underlying truth.

To demonstrate the ability of our method to distinguish between a cored ($\gamma = 0$) and cuspy ($\gamma = 1$) DM profile, in Figs. 2.3 and 2.4 we show the inferred posteriors on two test dwarf galaxies with the same DM density normalization ($\rho_0 = 10^7\,\mathrm{M}_\odot/\mathrm{kpc}^3$), scale radius ($r_s = 1\,\mathrm{kpc}$), and stellar profile, but with different inner density slopes ($\gamma = 0$ and $\gamma = 1$).

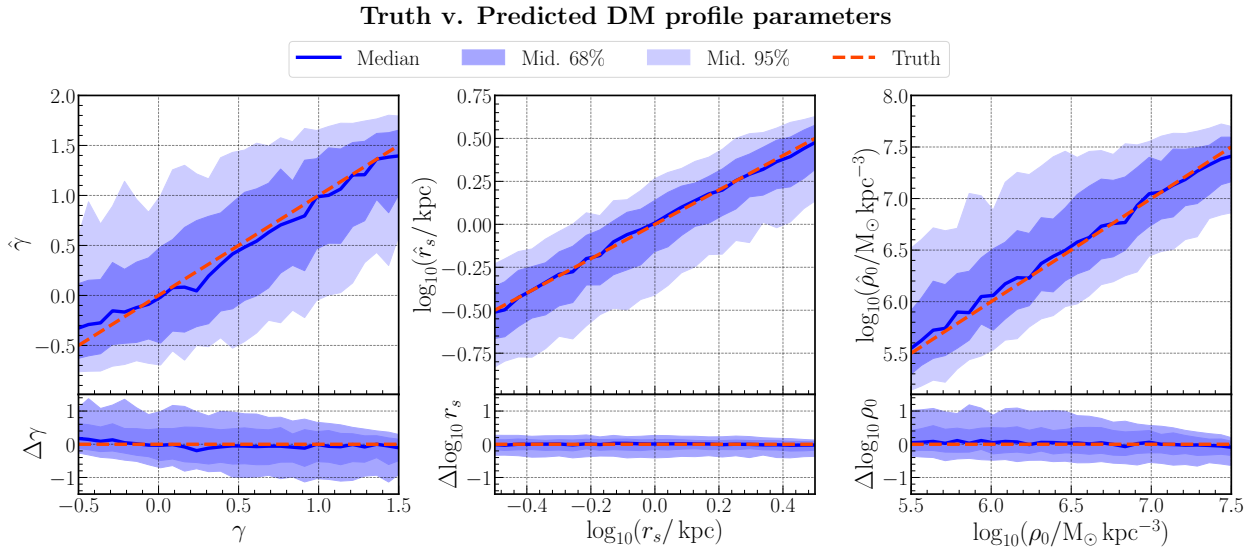

Figure 2.2: A comparison between the predicted and the truth values of the DM parameters on 10,000 test galaxies for our baseline scenario. For each galaxy, the predicted parameters are taken to be the marginal medians of the joint posterior binned in truth values. The median (solid blue line), middle-68% percentile (dark blue band), and middle-95% (light blue band) containment regions of each bin are shown. The diagonal dashed red line denotes where the predicted and truth values are equal. The bottom row shows the prediction error on the median $\Delta\theta \equiv \hat{\theta} - \theta_{\mathrm{truth}}$.

Figure 2.3 shows the posteriors on the recovered density profile (top row), enclosed mass profile (middle row), and orbital anisotropy profile (bottom row) as a function of halo-centric radius of a cored (left) and a cupsy (right) profile. The middle-68% and middle-95% credible intervals are shown as blue bands, and the truth profiles are shown with the dashed red lines. We find our method is able to successfully reconstruct the density, mass, and orbital anisotropy profiles at small and large radii.

We apply Jeans analysis on these two test galaxies using a procedure similar to that used in Ref. [254]. The posterior is approximated with nested sampling [316], [317] using the module DYNESTY [318] with $n_{\mathrm{live}} = 500$ live points and a convergence tolerance of $\Delta \ln \mathcal{Z} = 0.1$ on the estimated evidence. Details of the Jeans analysis procedure are provided in Appendix A.1.2. Figure 2.4 shows the joint and marginal DM posteriors from Jeans modeling (left panel) and our method (right panel) for $\gamma = 0$ (red) and $\gamma = 1$ (blue), with the middle-68% and middle-95% credible intervals as the contour lines. In the Jeans analysis, we see significant overlap between the γ posteriors of the cored and cuspy halos. On the other hand, γ posteriors inferred by our method have substantially smaller overlap. See Appendix A.2.2 for a quantitative comparison of the separability.

An important quantity in indirect searches for DM is the astrophysical J-factor—the inte-

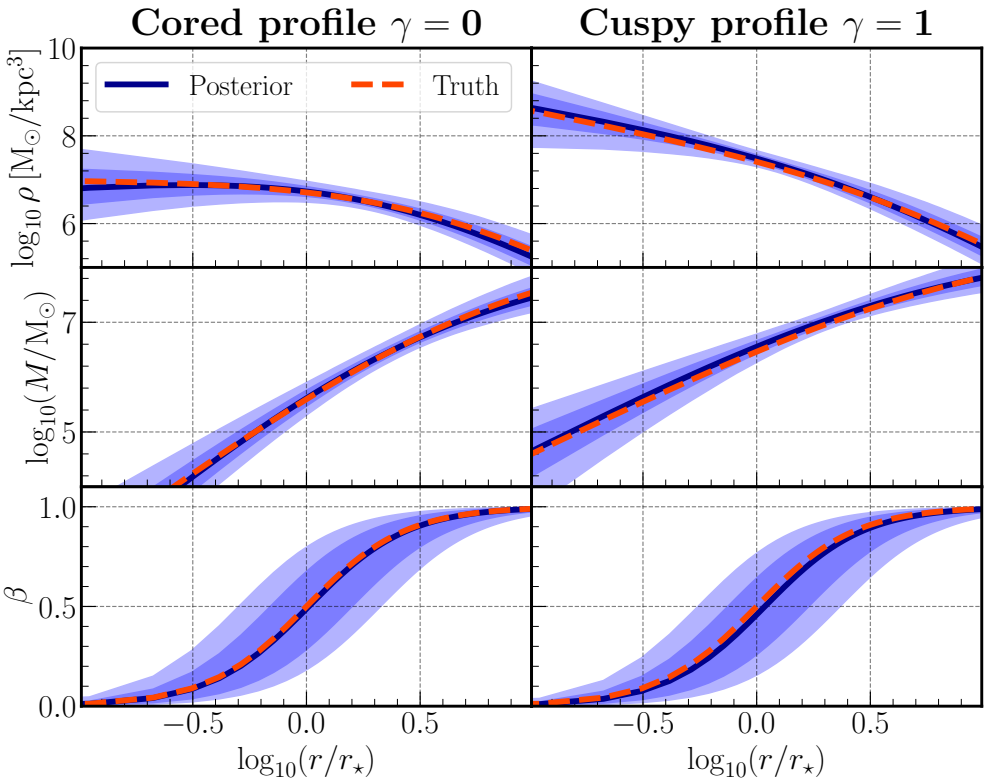

Figure 2.3: Example inferred posteriors of the density profile (top row), enclosed mass profile (middle row), and velocity anisotropy profile (bottom row) for dwarf galaxies with a cored DM profile (left) and a cuspy DM profile (right). The dashed red line is the truth profile, while the blue line and bands represent the median, middle-68%, and middle-95% credible intervals respectively.

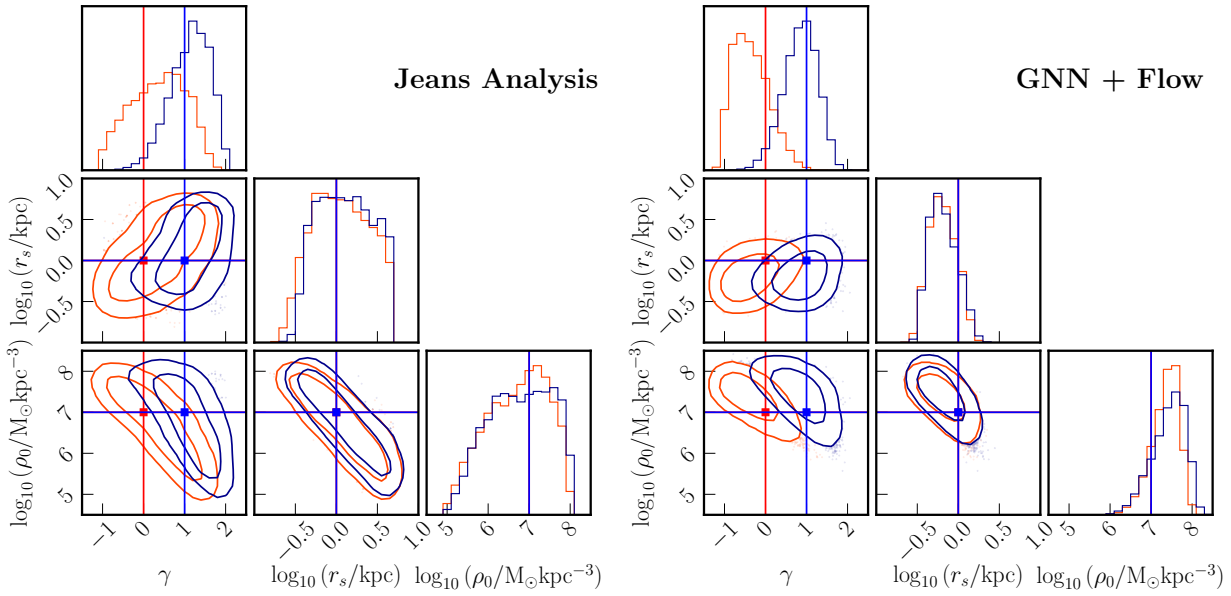

Figure 2.4: Example corner plots of the posterior DM parameters from Jeans dynamical modeling (left) and our method (right) on two test galaxies with cored DM profile (red) and cuspy DM profile (blue). Both galaxies have the same central slope ρ_0 and scale radius r_s. The contour lines show the 68% and the 95% credible intervals. It can be seen that our method provides a stronger constraint on the DM parameters and is able to distinguish more cleanly between a cored and cuspy profile.

gral along the line-of-sight s and over solid angle Ω of the squared DM density corresponding to a source target,

$$J = \int \mathrm{d}s \int \mathrm{d}\Omega \, \rho^2(s, \Omega), \tag{2.4}$$

accurately determining it is therefore important for robustly interpreting results of DM indirect detection experiments using dwarf galaxies as targets [88], [94]–[101].

In Figure 2.5, we show the inferred J-factors, normalized to the truth values, for 100 dwarf galaxies randomly sampled from our test dataset using our method (red error bars) and compare these to the corresponding J-factors obtained using Jeans analysis (green error bars). The middle panel of Figure 2.5 shows ratios of uncertainties on the $\log_{10} J$-factors obtained by our method to those obtained by the Jeans analysis. It can be seen that our method generically provides more stringent constraints than the Jeans analysis, as expected from the fact that the DM density parameters are overall better constrained. For the cases studied, this log-uncertainty is on average $\sim 20\%$ and up to a factor of ~ 2 smaller for our method than that obtained using Jeans analysis. The bottom panel shows the ratios of the absolute differences between the truth and predicted $\log_{10} J$-factors obtained by our method as compared to those obtained by the Jeans analysis. The scatter around unity indicates that

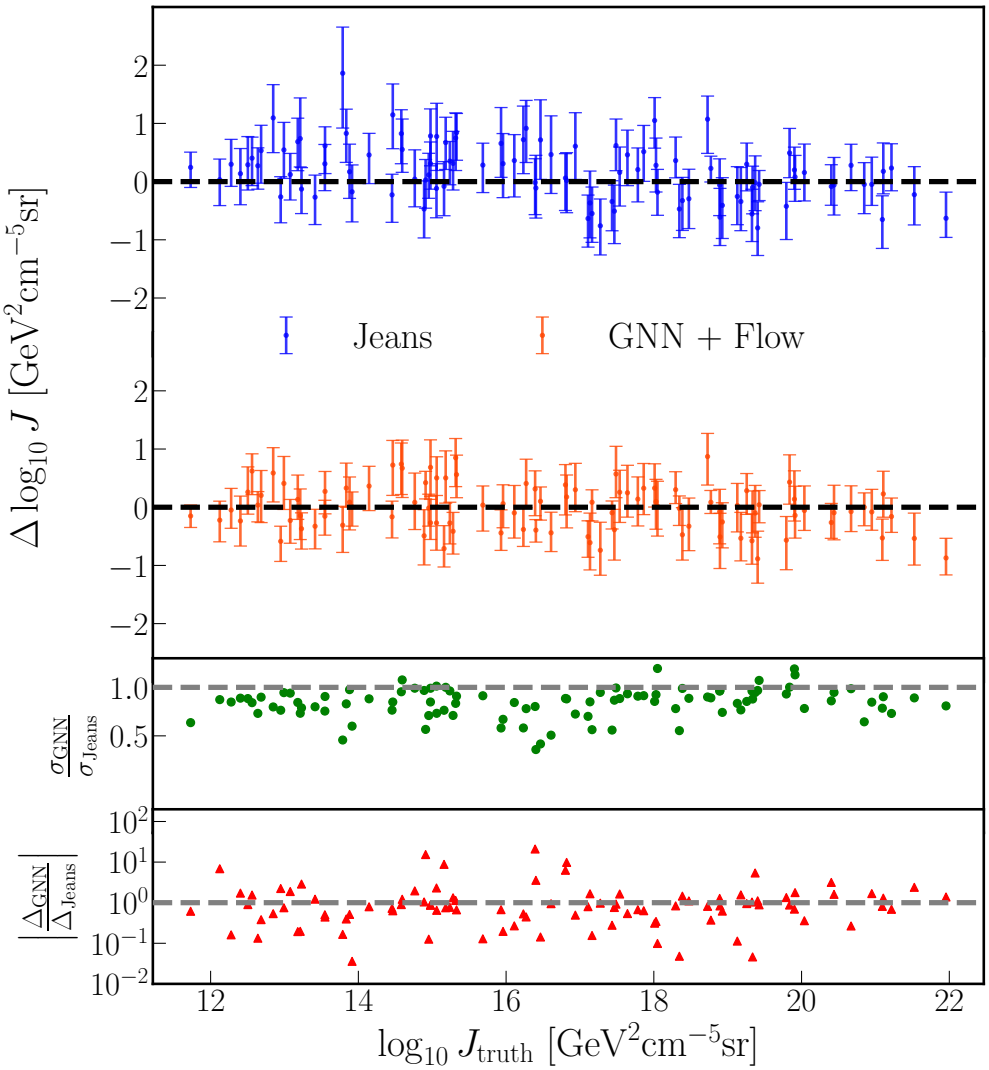

Figure 2.5: Comparison between the J-factors predicted by the Jeans analysis and our method. *Top*: Differences between the predicted and true $\log_{10} J$-factors for the Jeans analysis (blue) and our method (red). *Middle*: Ratios of the symmetrized errors on the $\log_{10} J$-factors, $\sigma_{\log_{10} J}$. *Bottom*: Ratios of the absolute differences between the truth and predicted $\log_{10} J$-factors. Our method can accurately recover the J-factor posteriors while generically providing tighter constraints than the traditional Jeans analysis.

neither method is systematically biased, as expected, and serves as an additional cross-check of our analysis. Details of the J-factor calculation and additional results comparing the two methods are shown in Appendices A.1.3 and A.2.3.

To conclude, this chapter introduces a novel method to reconstruct the DM density profiles of Milky Way dwarf galaxies from measured kinematics of tracer stars based on graph neural networks and simulation-based inference. The method compares favorably with and can outperform established approaches based on Jeans dynamical modeling in speed, flexibility, as well as constraining power. The latter is due to the fact that our method incorporates more information about the phase-space structure of bound stars, contrasted with Jeans-based methods which typically rely on the second moments of the velocity distribution. Additionally, the method simultaneously models the stellar light profile and does not require fitting it beforehand. Although in this chapter, we used simulations of orbitally-anisotropic spherical systems in order to enable a direct comparison with existing methods, a particular strength of our method is the ability to incorporate non-equilibrium dynamics using cosmologically-realistic simulations of isolated dwarfs as well as satellites of Milky Way-like systems, accounting for baryonic effects like tidal disruption [144] and supernova feedback [78]. We defer this extension, as well as application of our method to observational data, to future work.

Software and Data Availability

This research made use of the Bilby [186], dynesty [318], IPython [319], Jupyter [320], Matplotlib [321], nflows [322], NumPy [323], PyTorch [324], PyTorch Geometric [243], PyTorch Lightning [325], and SciPy [326] software packages.

Code used to reproduce the results of this chapter is available at https://github.com/trivnguyen/dsph_gnn.

Chapter 3

Generative Modeling for Halo Assembly Histories

Disclaimer

The content of this chapter was submitted to *Monthly Notices of the Royal Astronomical Society* and is currently under peer-reviewed. A copy is available on arXiv at https://arxiv.org/abs/2308.05145 [250]. The authors are Tri Nguyen, Chirag Modi, L.Y. Aaron Yung, and Rachel Somerville.

Abstract

The mass assembly history (MAH) of dark matter halos plays a crucial role in shaping the formation and evolution of galaxies. MAHs are used extensively in semi-analytic and empirical models of galaxy formation, yet current analytic methods to generate them are inaccurate and unable to capture their relationship with the halo internal structure and large-scale environment. This chapter introduces FLORAH, a machine-learning framework for generating assembly histories of ensembles of dark matter halos. We train FLORAH on the assembly histories from the GUREFT and VSMDPL N-body simulations and demonstrate its ability to recover key properties such as the time evolution of mass and concentration. We obtain similar results for the galaxy stellar mass versus halo mass relation and its residuals when we run the Santa Cruz semi-analytic model on FLORAH-generated assembly histories and halo formation histories extracted from an N-body simulation. We further show that FLORAH also reproduces the dependence of clustering on properties other than mass (assembly bias), which is not captured by other analytic methods. By combining multiple networks trained

on a suite of simulations with different redshift ranges and mass resolutions, we are able to construct accurate main progenitor branches (MPBs) with a wide dynamic mass range from $z = 0$ up to an ultra-high redshift $z \approx 20$, currently far beyond that of a single N-body simulation. FLORAH is the first step towards a machine learning-based framework for planting full merger trees; this will enable the exploration of different galaxy formation scenarios with great computational efficiency at unprecedented accuracy.

3.1 Introduction

In the Λ-Cold Dark Matter (ΛCDM) paradigm, dark matter (DM) halos form hierarchically and grow in mass through the mergers of smaller DM halos [39]. The assembly history of a halo or galaxy refers to the sequence of events that led to its formation and growth over cosmic time. It encompasses the accretion of matter, mergers with other halos or galaxies, and internal processes such as star formation and black hole growth. In simulations, this is often represented in the form of "merger trees" i.e. progenitor and descendant halos that are linked across cosmic time. In the modern paradigm of galaxy formation, the properties of galaxies (such as stellar mass and star formation rate) are believed to be closely linked to the assembly history of their halos and their formation environment. Understanding this intricate halo-galaxy connection remains one of the key open questions of modern astrophysics.

N-body simulations provide a powerful tool to directly follow the formation and evolution of DM halos as they merge with other halos and interact with the large-scale environment [149]. However, these simulations are computationally expensive, and their cost grows rapidly with the simulated volume and resolution. Therefore, they are often run with only dark matter and without including baryonic physics. It is not feasible to run a single simulation that can simultaneously capture the formation histories of halos from dwarf galaxies ($10^5 - 10^{10}$ $M_\odot$) to galaxy clusters ($10^{14} - 10^{15}$ $M_\odot$) up to high redshifts. It is currently challenging to run even a single dark matter-only simulation with a volume comparable to that of existing galaxy surveys and a mass resolution sufficient to accurately trace the merger histories of the halos hosting the galaxies that are detected in those surveys — let alone next-generation surveys. Computational limitations also hinder the exploration of different structure formation scenarios and cosmological parameters.

Semi-analytic models (SAMs) are commonly used to populate DM halo merger trees with galaxies. SAMs apply simplified prescriptions for baryonic physics (e.g., radiative cooling, star formation, supernova, and AGN feedback prescriptions, etc.) within cosmological merger trees to track the formation and evolution of galaxies and forward model their observable properties [276], [285], [327]–[330]. SAMs require significantly fewer computational

resources compared to N-body simulations and allow for exploration of a wider range of physical processes and parameters. They have been used to study many aspects of galaxy formation, ranging from large-scale clustering (e.g. [265]–[273]) to galaxy properties (e.g., colors, metallicities, sizes, cold gas contents) [274]–[281] to supermassive black hole formation and active galactic nuclei feedback [282]–[284] to reionization [285]–[287]. Similarly, semi-empirical models such as UniverseMachine and EMERGE link galaxy observables to halo formation histories [331]–[333]. Accounting for the dependence of halo clustering on properties other than mass (assembly bias) has been shown to be important for accurately interpreting galaxy clustering measurements [e.g 269], [271], [334]. "Decorated" halo occupation distribution models have been proposed, which again require knowledge of the halo formation history and/or structural properties [e.g. 258].

To overcome these computational challenges, previous works have developed analytic methods that rely on the Extended Press-Schechter (EPS) formalism [335], [336]. EPS trees are constructed by sampling the conditional mass probability $p(M_1|M_0, z_0, z_1)$ that a halo with mass M_0 at redshift z_0 had a mass of M_1 at an earlier redshift $z_1 > z_0$. For halos of any given initial mass M_0 and redshift z_0, the algorithm uses Monte-Carlo methods and assumes the Markov property to construct its past merger histories [e.g. 337]–[343]. However, this approach has many limitations. The resulting MAHs and merger trees are known to disagree with the results of N-body simulations at the factor of few levels [344]–[346]. Moreover, EPS-based techniques are fundamentally unable to capture the relationship between the assembly history, the halo structure, and the environment. For a comparison between different EPS-based techniques for generating merger trees, we refer readers to [347].

A recent study has applied Generative Adversarial Networks and Convolutional Neural Networks to generate ensembles of merger trees [348]. Although this approach seems promising, this work did not incorporate the correlation between halo formation history and environment and did not show a complete set of diagnostic tests on the resulting merger tree ensembles. Another recent study using differential programming was able to accurately capture both the MAHs and assembly bias [349]. However, this approach smooths out the MAH of individual halos, making it difficult to capture merger events. It is also unclear how the halo structure and formation environment can potentially be included in the framework.

In this chapter, we introduce florah (FLOw-based Recurrent model for Assembly Histories) a deep generative model based on recurrent neural network and normalizing flows, to generate assembly histories of halos. As a first step, we focus on generating the mass assembly histories (MAHs) and DM concentration histories of only the main progenitor branches (MPBs). The MPB tracks the most massive progenitor of a halo and thus is the most important for understanding the assembly history. MPBs can be naturally modeled as

a time-ordered sequence; hence we use a recurrent neural network to learn their representative features. Furthermore, for a halo of given mass and concentration, we are interested in learning the full distribution of possible assembly histories and hence we combine this recurrent network with a normalizing flow to learn this distribution. Once trained, FLORAH can be used to generate the MAHs and DM concentration histories of MPBs consistent with N-body simulations. Although FLORAH only generates the MPBs of merger trees, it is the first framework to build in the correlation between halo structural properties and merger history.

To demonstrate this, we use the Santa Cruz SAM (SC-SAM) [276], [350], with free parameters calibrated to reproduce a subset of observed galaxy properties at $z = 0$ (see [271] for detailed calibration criteria), to populate FLORAH-generated MPBs with galaxies and show that FLORAH can correctly capture the assembly bias and environmental dependence of galaxies, which existing EPS-based methods cannot reproduce. In addition, by combining multiple networks trained on different simulations with varying resolutions and redshift ranges, we are able to generate high-resolution MPBs to ultra-high redshifts, from $z = 0$ to $z = 20$. This allows us to overcome the dynamic range limit of a single simulation.

This chapter is organized as follows. Section 3.2 provides a detailed description of the training simulation and the algorithms employed to extract the merger trees. Section 3.3 outlines our proposed method, including the data preprocessing procedure for a single simulation (Section 3.3.1) and for the combination of multiple simulations (Section 3.3.1). We describe the neural network architecture and optimization in Section 3.3.2, followed by an explanation of the generation procedure in Section 3.3.3. In Section 3.4, we show the generation results with VSMDPL (Section 3.4.1) and GUREFT (Section 3.4.2), while Section 3.4.3 presents a comparison with the combined GUREFT and VSMDPL simulations. We compare FLORAH to previous approaches for modeling MAHs, and discuss the limitations of the current approach and potential avenues for future research in Section 3.5. We conclude the chapter in Section 3.6.

3.2 Simulation

Table 3.1: The simulation specifications of VSMDPL and the four GUREFT simulations. The third and second-to-last columns show the minimum and maximum root mass of the training dataset for each simulation. Note that the four GUREFT simulations are combined into one training dataset. The last column shows the minimum progenitor mass during the generation process.

	Box size	M_{DM}	N	ϵ	n_{snap}	Redshift range	min $M_{\mathrm{vir},0}$	max $M_{\mathrm{vir},0}$	$M_{\mathrm{min,halo}}$
	$[\mathrm{Mpc\,h^{-1}}]$	$[\mathrm{M_\odot\,h^{-1}}]$		$[\mathrm{kpc\,h^{-1}}]$			$[\mathrm{M_\odot\,h^{-1}}]$	$[\mathrm{M_\odot\,h^{-1}}]$	$[\mathrm{M_\odot\,h^{-1}}]$
GUREFT-05	5	9.92×10^3	1024^3	0.16	171	$5.89 - 40.0$	4.96×10^6	2.14×10^9	9.92×10^5
GUREFT-15	15	2.68×10^5	1024^3	0.49	171	$5.89 - 40.0$	2.14×10^9	2.14×10^{10}	2.68×10^7
GUREFT-35	35	3.40×10^6	1024^3	1.14	171	$5.89 - 40.0$	2.14×10^{10}	2.14×10^{11}	3.40×10^8
GUREFT-90	90	5.78×10^7	1024^3	2.93	171	$5.89 - 40.0$	2.14×10^{11}	n/a	5.78×10^9
VSMDPL	160	6.20×10^6	3840^3	$1.0 \rightarrow 2.0$	151	$0.00 - 24.9$	3.10×10^8	n/a	6.20×10^8

For training datasets, we extract merger trees from the GUREFT (Gadget at Ultra-high Redshift with Extra Fine Timesteps) [154] suites and the VSMDPL simulation from the MultiDark suite [153]. Both GUREFT and VSMDPL are DM-only, N-body simulations run with the GADGET-2 smoothed particle hydrodynamics code [150], [151]. The cosmological parameters are broadly consistent with the Planck 2013 results [31], namely $\Omega_m = 0.307, \Omega_\Lambda = 0.693, h = 0.678, \sigma_8 = 0.823, n_s = 0.960$. The specifications of GUREFT and VSMDPL are summarized in Table 3.1.

The GUREFT suite consists of four boxes with sizes of 5, 15, 35, and 90 $\mathrm{Mpc\,h^{-1}}$. Each box has the same number of particles 1024^3, with the smaller box having a higher mass resolution. Therefore, each box captures the assembly histories of a different halo mass range. The GUREFT boxes have 171 snapshots from $z \approx 40$ to 6 at a high temporal resolution, making them the ideal simulations to capture the merger rates and assembly histories at high to ultra-high redshift. For $z \lesssim 6$, we instead use the VSMDPL simulation, which has 151 snapshots from $z = 0$ to 25. Thanks to its high mass resolution, with a particle mass of 6.2×10^6 $\mathrm{M_\odot\,h^{-1}}$, VSMDPL has a higher number of well-resolved and low-mass halos at high redshift. This allows us to combine VSMDPL with GUREFT into one training dataset spanning a wider redshift and mass range. We will describe this procedure in more detail in Section 3.3.1.

We construct merger trees for each GUREFT simulation and use the publicly available merger trees from VSMDPL. All merger trees used in this work are constructed using the ROCKSTAR halo finder and the CONSISTENTTREE algorithm [351], [352]. Note that merger trees may be sensitive to the details of the halo finding and tree construction algorithms. We plan to experiment with different parameters and tree reconstruction algorithms in future work.

Figure 3.1 displays the median MAHs on the main branches of GUREFT and VSMDPL. The overlapping MAHs between the two simulations from $z \approx 13-5$ are crucial for combining them. We have only shown the median MAHs down to approximately $100\,M_{\rm DM}$, where $M_{\rm DM}$ denotes the mass of the DM particles in the corresponding simulations. Halos with masses below this threshold cannot be reliably identified by ROCKSTAR. Therefore, this limits the maximum redshift that we can generate for each simulation, with a maximum of $z \approx 20$ (from GUREFT-05), as shown in Figure 3.1. Furthermore, the less massive root halos of VSMDPL ($\lesssim 10^{10}\,\mathrm{M_\odot}$) do not overlap with GUREFT, which imposes a further limit on the maximum generation redshift for these halos. Additionally, each simulation has a unique particle mass and a resolved mass limit, which can complicate the process of combining the simulations. We will provide a more detailed discussion of the effects of particle mass and resolution limitations in Section 3.3.1.

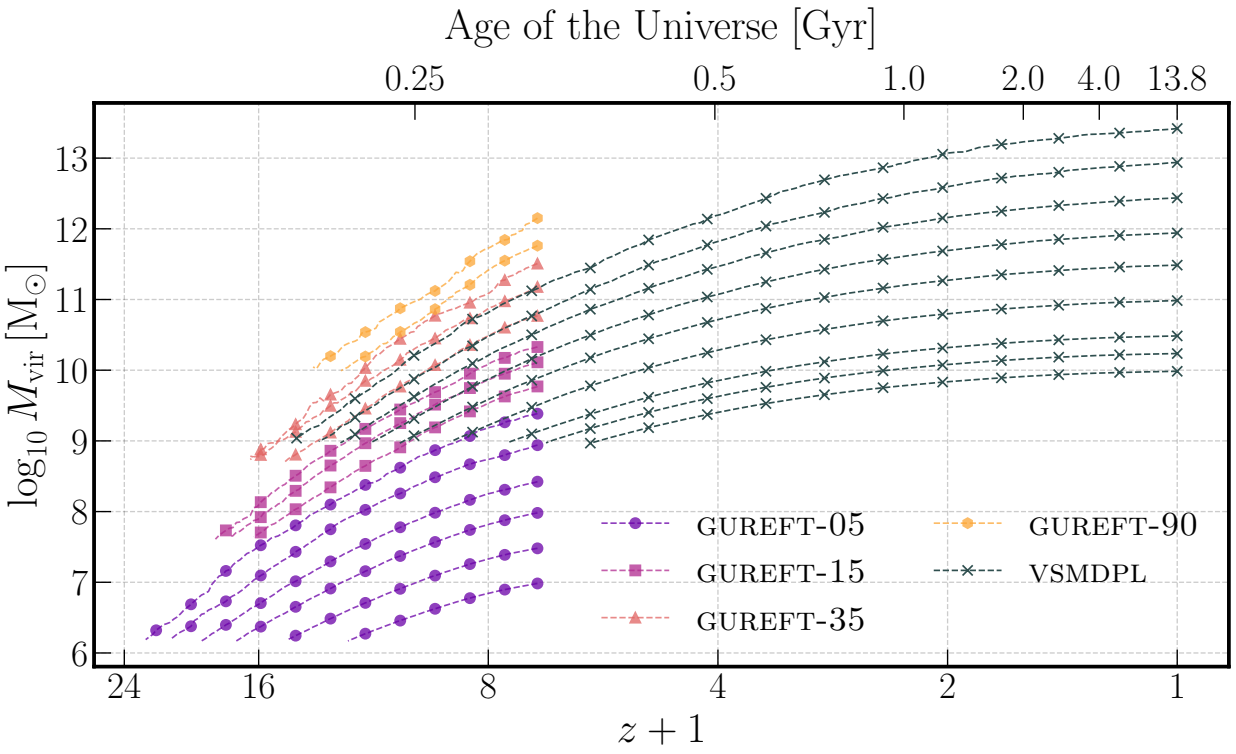

Figure 3.1: The median MAHs of the main progenitor branches of GUREFT and VSMDPL down to $100\,M_{\rm DM}$, where $M_{\rm DM}$ is the mass of the DM particle in the corresponding simulations.

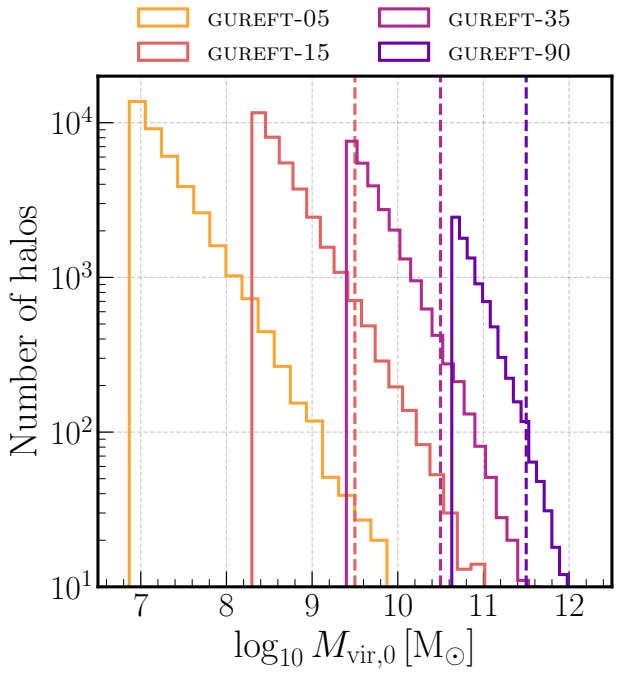

Figure 3.2: Distributions of M_{vir} for four GUREFT boxes at $z = 5.89$. Each dashed vertical line represents the additional minimum mass cut (on top of the 500 M_{DM} cut) in Section 3.3.1 for the corresponding box in the combined box scheme. We do not impose any additional minimum mass cut on GUREFT-05.

3.3 Methodology

In this section, we outline the data preprocessing, neural network architecture, and the optimization and generation process used in this work. A schematic illustration of FLORAH is displayed in Figure 3.3. As a first approach, we aim to generate only the MPBs of merger trees since they contribute a first-order impact on the assembly histories of halos. We plan to extend our framework to generate secondary branches and other branches in future work.

3.3.1 Data preprocessing

Using a single simulation

In this section, we describe the data preprocessing steps for a single simulation. The same steps apply for the four GUREFT boxes and VSMDPL with different parameters because they have different redshift ranges and resolutions. We will highlight the parameters that differ in Section 3.4, where we describe the training dataset in more detail. The following steps are the same for all simulations.

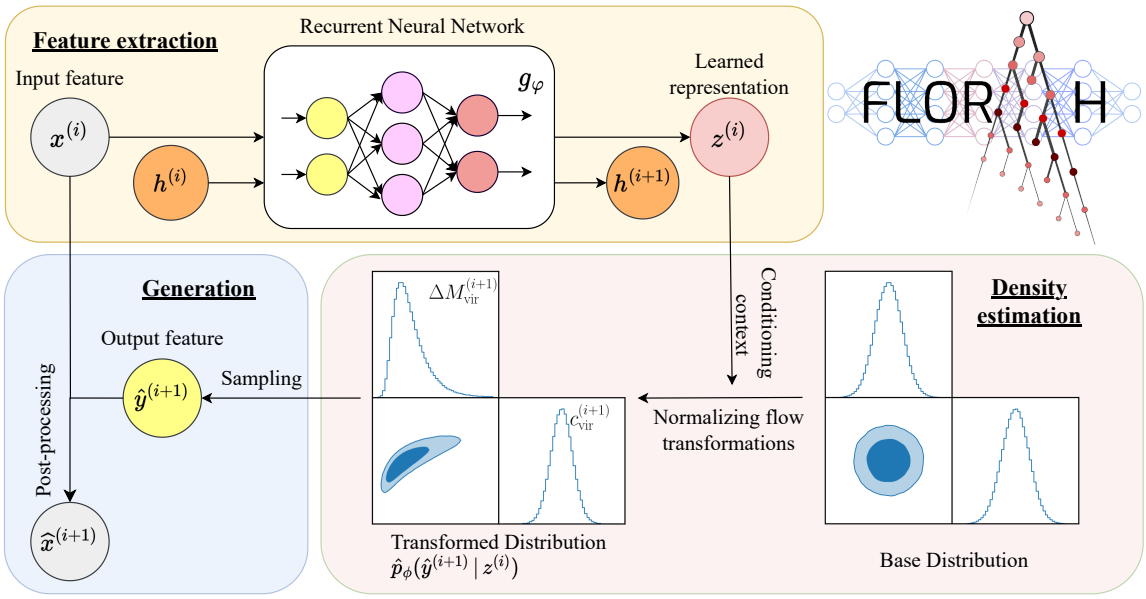

Figure 3.3: A schematic illustration of FLORAH.

As a first step, we extract only the main progenitor branches (MPBs) and exclude all branches with root halos containing fewer than 500 DM particles; halos below this limit may have poorly resolved progenitors. As previously mentioned in Section 3.2, the lower limit of the progenitor mass is about $100\,M_{\rm DM}$, below which ROCKSTAR may not reliably identify halos. We remove these unresolved halos during the post-processing of the generated MPBs (described in Section 3.3.3), rather than during the training dataset creation, as we have found that it leads to improved model performance. This resolution limit is particularly important when combining multiple simulations with varying $M_{\rm DM}$ into a single training dataset, which we further elaborate on in Section 3.3.1. In addition, MPBs from CONSISTENTTREE may sometimes have progenitor masses greater than their descendants. This is a common phenomenon that occurs due to the tree construction algorithm (including CONSISTENTTREE) misidentifying the halos and their mass assignments. As with the unresolved halos, we will correct for this effect during the generation and post-processing of the MPBs in Section 3.3.3, rather than during the training phase.

To enhance the generalization ability of our model, we employ data augmentation by creating multiple "sub-branches" for each MPB. We begin by selecting the initial snapshot from a uniform distribution up to the first 50 snapshots ($z \approx 2$ for VSMDPL and $z \approx 10.6$ for GUREFT). Next, we subsample the MPB by selecting every 2 to 6 snapshots (chosen uniformly) up to either a maximum redshift of $z_{\rm max,train}$ or a maximum length of 20 halos. We limit the length of each sub-branch" to 20 halos to prevent overly long sequences that could impact the learning process of the recurrent neural network [353], [354]. Subsampling

the tree in this way has a few advantages. First, it allows us to cover a wider range of redshifts, improving the overall generalization of the model. Second, it has a smoothing effect on the MAHs. Third, randomizing the time steps helps prevent the model from becoming overly reliant on one particular set of time steps and improves generalization. We repeat the above data augmentation process 15 times for each tree.

The input features to the recurrent neural network at each stage are the logarithm of the virial mass $\log_{10} M_{\rm vir}$, the Navarro-Frenk-White DM concentration $c_{\rm vir}$ [355], and scale factors a of the halo and the next progenitor. The concentration is not given directly by ROCKSTAR and instead computed from the relation $c_{\rm vir} \equiv r_{\rm vir}/r_s$ where r_s is the Navarro-Frenk-White (NFW) scale radius [355] and $r_{\rm vir}$ is the virial radius. Note that the NFW scale radius in ROCKSTAR is computed in two different ways, by fitting the density profiles or by converting from the radius of the maximum circular velocity [152]. Here we use the value provided by fit to the density profiles. If a halo undergoes a recent major merger or has multiple density peaks, the fitting procedure in ROCKSTAR can fail, and r_s will be capped at $r_{\rm vir}$.This can create a "pile up" in the concentration for halos with $r_s \approx r_{\rm vir}$. Therefore, we require that $(r_s - r_{\rm vir})/r_s > 0.1$, or equivalently $c_{\rm vir} > 1.1$. We remove all halos that do not pass this cut during the generation and post-processing of the MPBs in Section 3.3.3, as with the unresolved halos. Including $c_{\rm vir}$ helps the model learn assembly histories more accurately because $c_{\rm vir}$ has been found to capture the environmental dependency in many assembly bias studies [356]–[359]. We will experiment with expanding to more halo features (e.g. spin, shape) and environment features (e.g. local density, number of neighboring halos) in future work. The scale factors of the halo and its next progenitor serve as the time features in our framework. We have experimented with other time features like redshift and age and found similar performance.

For the output of the network, we model the logarithm of the *change in* mass, defined as:

$$\Delta \log_{10} M_{\rm vir}^{(i+1)} = \log_{10}(M_{\rm vir}^{(i+1)}/M_{\rm vir}^{(i)}), \tag{3.1}$$

for the $(i+1)$-th halo, and the concentration $c_{\rm vir}$ of the next progenitor. We found that using accreted masses ($\Delta \log_{10} M_{\rm vir}^{(i+1)}$) as targets, instead of progenitor masses ($\log_{10} M_{\rm vir}^{(i+1)}$), improves the model performance. During generation, progenitor masses can be derived from masses and accreted masses using Equation 3.1.

Combining multiple simulations

We present a procedure for combining the four GUREFT boxes (GUREFT-05, GUREFT-15, GUREFT-35, and GUREFT-90) into a single training dataset, which we denote as GUREFT-

C. The four GUREFT boxes exhibit significant overlap in halo masses as depicted by the distributions of $M_{\rm vir}$ in Figure 3.2. However, due to differences in resolutions and volumes, MAH for halos in overlapping mass bins could vary by a significant amount, particularly at high redshifts. In such instances, the box with the highest resolution (i.e., the one with the lowest $M_{\rm DM}$) would yield a more accurate MAH and we would like to only use these for training. Hence to combine the GUREFT boxes, we impose an additional minimum root mass cut of $\log_{10} M_{\rm min,root} \approx 9.5$, 10.5, 11.5 dex for GUREFT-15, GUREFT-35, and GUREFT-90 respectively. These cuts are applied directly to the root halos at $z = 5.89$ and before any data augmentation step. These thresholds are chosen to strike a good balance between MAH accuracy (which increases for smaller boxes at a given mass) and the number of halos in a mass bin (which is lower for smaller boxes), as having too few halos can adversely affect the training process. No additional cut is applied to GUREFT-05, which has the highest resolution, resulting in an over-representation of low-mass halos in our training dataset. Despite the imbalance in the number of low and high-mass halos, we show in Section 3.4 that we are able to capture the high-mass halos accurately. The remaining data preprocessing steps follow the guidelines outlined in Section 3.3.1. Each GUREFT box is preprocessed separately, and the resulting datasets are combined into a single, large training dataset denoted as GUREFT-C.

3.3.2 Neural network architecture and optimization

We model each sub-branch as a sequence of N halos, with the root halo denoted with the index zero. The input and target feature vectors are:

$$\vec{x} = \{x^{(i)} \in \mathbb{R}^{f_{\rm in}}\} = \{\log_{10} M_{\rm vir}^{(i)}, c_{\rm vir}^{(i)}, a^{(i)}, a^{(i+1)}\}, \tag{3.2}$$

$$\vec{y} = \{y^{(i+1)} \in \mathbb{R}^{f_{\rm out}}\} = \{\Delta \log_{10} M_{\rm vir}^{(i+1)}, c_{\rm vir}^{(i+1)}\}, \tag{3.3}$$

respectively, where $f_{\rm in} = 4$, $f_{\rm out} = 2$, and $i = 0, ..., N-2$. We do not include any "end-of-sequence" token in our framework, so the feature vector includes only the first $N-1$ halos. During the generation process (Section 3.3.3), we can choose to terminate the assembly history at a maximum redshift or minimum progenitor mass. Our goal is to learn the true conditional distribution of $y^{(i+1)}$, denoted as $p(y^{(i+1)}|\{x^{(\leq i)}\})$.

We use a recurrent neural network $g_\varphi : \mathbb{R}^{N_{\rm in}} \to \mathbb{R}^H$ with parameters φ to extract H

summary features from the input features of each halo. The summary features are then:

$$\vec{z} = \{z^{(i)} \in \mathbb{R}^H\} = \{g_\varphi(x^{(i)}, h^{(i)})\}, \tag{3.4}$$

$$h^{(i)} = z^{(i-1)} \text{ if } i > 0 \text{ else } 0. \tag{3.5}$$

The hidden state $h^{(i)}$ is dependent on the input features of all the previous halos $x^{(\leq i)}$, which allows the network to "memorize" the entire assembly history. Our recurrent network consists of 4 Gated Recurrent Unit (GRU) layers, each with $H = 128$ hidden channels. For the recurrent layers, we have also experimented with different network architectures like Transformers [360], adding a decay mechanism to the hidden states [361], and adding time-embedding layers. We found that these choices result in similar performance while being more computationally intensive than our current model.

To learn the true probability distribution $p(y^{(i+1)} \mid x^{(\leq i)})$, we use a normalizing flow [232], [309], [362] that is conditioned on the summary features $z^{(i)}$. The flow thus estimates a conditional probability distribution $\hat{p}_\phi(y^{(i+1)} \mid g_\varphi(x^{(i)}, h^{(i)}))$ with learnable parameters ϕ. Our normalizing flow model consists of 4 Masked Autoregressive Flow (MAF) transformations. Each MAF includes a 4-layer Masked Autoencoder for Distribution Estimation (MADE) with a hidden dimension of 128 [310], [362].

During training, we optimize the parameters $\{\varphi, \phi\}$ of the recurrent network and flow simultaneously using the negative log-density

$$\mathcal{L} = -\sum_{i=0}^{N-1} \log \hat{p}_\phi(y^{(i+1)} \mid g_\varphi(x^{(i)}, h^{(i)})), \tag{3.6}$$

as the optimization loss. We use the AdamW optimizer [314], [315] with parameters $(\gamma, \beta_1, \beta_2, \lambda) = (0.001, 0.9, 0.98, 0.01)$, where γ is the learning rate, β_1 and β_2 are the running average coefficients, and λ is the weight decay coefficient. At the end of each *epoch* (defined as one full iteration over the training set), we evaluate the loss on the validation samples and reduce the learning rate by a factor of 10 if no improvement is seen after 20 epochs. We terminate the training process if the validation has not improved after 40 epochs. The training process terminated at ~ 200 epochs or ~ 2 hours on a single NVIDIA Tesla V100 GPU.

3.3.3 Generation and post-processing

Once the model is trained, the generation process for one tree is as follows:

1. Select an initial halo feature $x^{(0)}$, which is the initial mass $\log_{10} M_{\rm vir}^{(0)}$ and concentration $c_{\rm vir}^{(0)}$.

2. Select a list of scale factors $\{a^{(i)}\}$ as time steps. FLORAH is robust across various time steps, as long as the chosen time steps remain within a reasonable range relative to the training time steps.

3. Starting with the first time step $i = 0$, pass the halo feature $x^{(i)}$ and the hidden state $h^{(i)}$ (with $h^{(0)}$ initialized to 0) through the GRU layers and extract the summary features $z^{(i)} = g_\varphi(x^{(i)}, h^{(i)})$.

4. Sample $\hat{y}^{(i+1)} \sim \hat{p}_\phi(\hat{y}^{(i+1)} \,|\, z^{(i)})$ using the normalizing flow layers. This returns the accreted mass $\Delta \log_{10} M_{\rm vir}^{(i+1)}$ and the DM concentration of the progenitor halo $c_{\rm vir}^{(i+1)}$[1].

5. Convert the accreted mass to progenitor mass using Equation 3.1 and set the feature of the progenitor halo to be $x^{(i+1)} = (\log_{10} M_{\rm vir}^{(i+1)}, c_{\rm vir}^{(i+1)}, a^{(i+1)}, a^{(i+2)})$.

6. Update the time step to $i + 1$ and repeat Steps (ii) to (v) until a minimum progenitor mass $M_{\rm min,halo}$ or a maximum redshift $z_{\rm max,gen}$. We elaborate on the procedure to choose these thresholds below.

As mentioned in Section 3.3.1, halos with fewer than 100 DM particles are not reliably identified by ROCKSTAR. For a model trained on a single simulation (i.e. VSMDPL in our case), we set the minimum progenitor mass to be $M_{\rm min,halo} = 100\, M_{\rm DM}$, where $M_{\rm DM}$ is the DM particle mass of the simulation. For a model trained on a combined simulation (i.e. GUREFT-C in our case), due to the additional mass cut described in Section 3.3.1, the generated MPBs will have varying mass resolutions depending on their root masses. In the case of GUREFT-C, we determine the DM particle mass $M_{\rm DM}$ based on the root mass of the MPB at $z = 5.89$ and set $M_{\rm min,halo} = 100\, M_{\rm DM}$ as before. The minimum progenitor masses are shown in the last column of Table 3.1 for each simulation. As for the maximum generation redshift $z_{\rm max,gen}$, we generally recommend $z_{\rm max,gen}$ to be smaller than $z_{\rm max,train}$ to avoid extrapolation. However, we will show in Section 3.4 that FLORAH can extrapolate the MPBs beyond $z_{\rm max,train}$.

Throughout the formation process of an MPB, the mass of the halo increases monotonically. However, as mentioned in Section 3.3.1, due to the misidentification of halos and mass assignments, progenitor halos may sometimes be assigned a mass greater than that of their descendant halos. For FLORAH-generated MPBs, this can be corrected by simply re-sampling the mass and concentration in Step (iv). However, for MPBs in N-body simulations, the specific corrective measures will depend on the nature of the problem and the characteristics

[1] As mentioned in Section 3.3.1, ROCKSTAR caps r_s at $r_{\rm vir}$, resulting in a "pile up" at around $c \approx 1$. To remove this population, we require the generated concentration $c_{\rm vir}^{(i+1)} > 1.1$ and simply re-sample the halo if this condition is not satisfied.

of the simulation data. Thus, to compare FLORAH-generated MPBs with N-body MPBs, we adopt the following simple method. For an MPB with N halos, starting from the last progenitor halo at time step $i = N - 1$ (with the root halo at time step 0), we compare its mass $M_{\rm vir}^{(i)}$ with the mass of the next halo $M_{\rm vir}^{(i-1)}$. If $M_{\rm vir}^{(i)} > M_{\rm vir}^{(i-1)}$, we assume that no accretion occurs and set the mass and concentration of the next halo $(M_{\rm vir}^{(i-1)}, c_{\rm vir}^{(i-1)})$ to be the progenitor's, i.e. $(M_{\rm vir}^{(i)}, c_{\rm vir}^{(i)})$. We then move forward to the next time step $i-1$ until we reach the root halo at time step 0.

3.4 Results

In Section 3.4.1, we train FLORAH on VSMDPL to show results for learning MPBs of a single simulation. Then we train FLORAH on the combined GUREFT-C simulations to show results for combining multiple simulations at different resolutions. Finally, in Section 3.4.3, we combine the trained models to extend the mass and redshift coverage of FLORAH.

3.4.1 Learning the VSMDPL simulation

We trained FLORAH on VSMDPL to test our method on its capability to learn the assembly histories of a single simulation. We refer to this model as FLORAH-V. For our training and validation datasets, we extracted 306,014 and 34,400 MPBs, respectively, from a $(80\,{\rm Mpc\,h^{-1}})^3$ sub-volume of VSMDPL and applied the preprocessing step in Section 3.3.1. The maximum training redshift is set to be $z_{\rm max,train} = 10$. The training process terminated at ~ 200 epochs or ~ 2 hours on a single NVIDIA Tesla V100 GPU.

For our test dataset, we extracted 387,031 MPBs from a *different* $(80\,{\rm Mpc\,h^{-1}})^3$ VSMDPL sub-volume. For each MPB in the test dataset, we used the initial halo at $z = 0$, taking its mass and concentration as the initial input feature $x^{(0)} = (\log_{10} M_{\rm vir}^{(0)}, c_{\rm vir}^{(0)})$ for the generation process. We sampled the scale factors every 2 – 6 snapshots[2] starting from $z = 0$ to obtain the time features of each branch $a^{(i)}$. We also sampled up to a redshift of $z_{\rm max,gen} = 14$, beyond the maximum training redshift ($z_{\rm max,train} = 10$), to explore the model's extrapolation capability. As the length of each MPB is set by choosing $a^{(i)}$, which are chosen randomly, we can generate MPBs with variable lengths, and the upper bound in the number of progenitor halos depends on the minimum mass. We generated 387,031 FLORAH-V MPBs to match the number obtained from VSMDPL.

[2]This choice is done out of convenience to compare the generated MPBs with the N-body simulation. In general, the time steps can be set randomly as long as they are within a reasonable range of the training time steps, as mentioned in Step (ii) of Section 3.3.3.

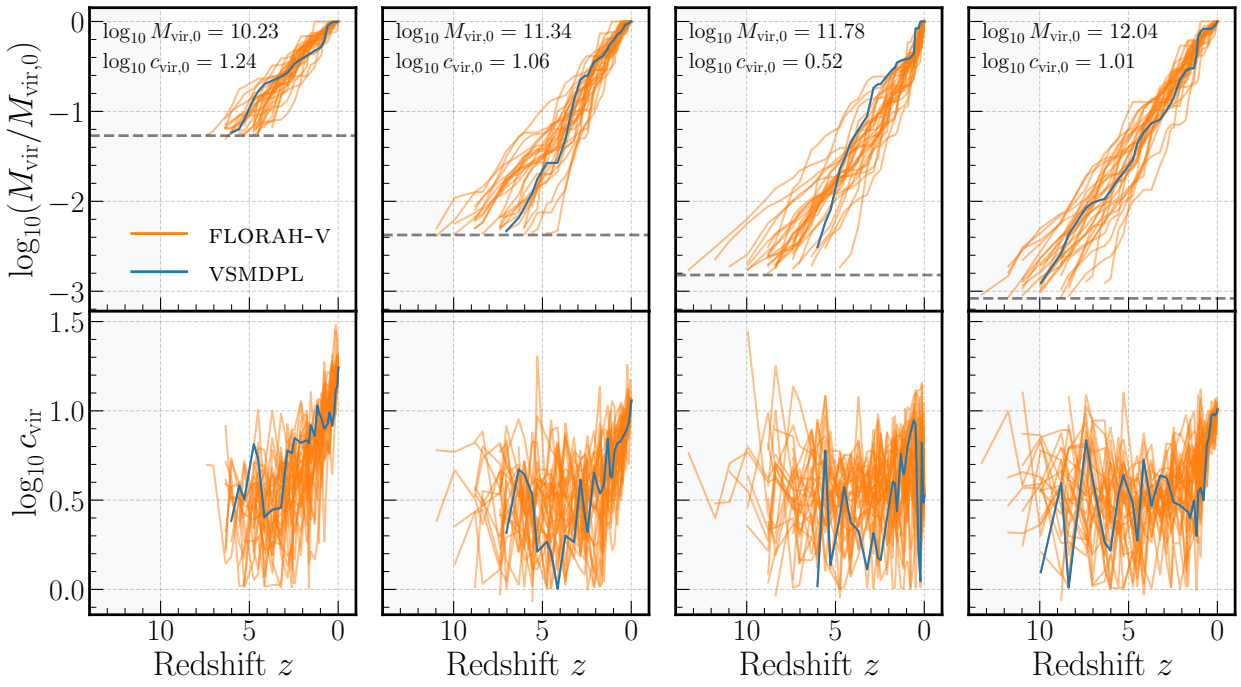

Figure 3.4: Example mass and concentration assembly histories generated by FLORAH-V. Each column shows the mass (top) and concentration (bottom) assembly histories of an example root halo from the VSMDPL simulation (blue) and 30 different realizations by FLORAH-V. The shaded gray box denotes the "extrapolation region" as we only train FLORAH-V up to a maximum training redshift of $z = 10$.

In Figure 3.4, we show an example of a few histories generated by FLORAH-V. Each column shows the mass (top panel) and concentration (bottom panel) histories of an example root halo in the VSMDPL simulation (blue). For each root halo, we use FLORAH-V to generate 30 different realizations of the histories (orange). The shaded regions denote the extrapolation region beyond $z_{\rm max,train} = 10$, as mentioned above. This figure demonstrates the stochasticity of the assembly histories of halos: for the same descendant mass and concentration, we can arrive at vastly assembly scenarios. In addition, we see that FLORAH-generated MAHs can capture both smooth accretion, characterized by a steady increase in the MAHs, as well as implicit merger events, characterized by sharp jumps in the MAHs. This highlights the advantage of FLORAH over other approaches, distinguishing it from other methods that predominantly capture the average trends of MAH.

Progenitor-descendant ratio distribution

In the following sections, we present various tests to validate the mass and concentration histories generated by FLORAH-V. In Figure 3.5, we show the distribution of the progenitor-descendant mass ratios $\log_{10} M_P/M_D$ across four mass bins (columns) and three redshift slices (rows). The distribution of $\log_{10} M_P/M_D$ is learned directly by FLORAH-V during the training process and thus presents a good first validation test. We see good agreement between the distributions from VSMDPL (shaded blue) and FLORAH-V (solid orange). Note that because the last mass bin with $\log_{10} M_{\rm vir,0} \in [13.0, 14.0]$ dex contains only 373 MPBs, the distributions are noticeably noisier. In each bin of the first row of Figure 3.5, the descendant halos are also the root halos and thus have roughly the same virial mass. However, we see that the distribution of the progenitor-descendant mass ratios can span a wide range of values, which further demonstrates the stochasticity of the MAHs between redshifts.

Mass and DM concentration histories

In Figure 3.6, we show the median, middle-68% percentile, and middle-95% percentile containment region of the assembly histories of VSMDPL (blue) and FLORAH (orange), along with their residuals, in four mass bins. We normalize the MAHs by dividing the progenitor masses by the root mass so that they can be more easily compared between bins. The residual of the containment region is computed by averaging the residuals of the corresponding upper and lower percentile curves. As can be seen from the slopes of the MAHs, low-mass halos tend to form earlier than high-mass halos, as has been observed in previous studies of galaxy formation (e.g. [334]). The MAHs plateau out at around the resolution limit $M_{\rm min,halo} = 100\, M_{\rm DM}$ (dashed black line, computed from the high end of each mass bin), as expected, because we

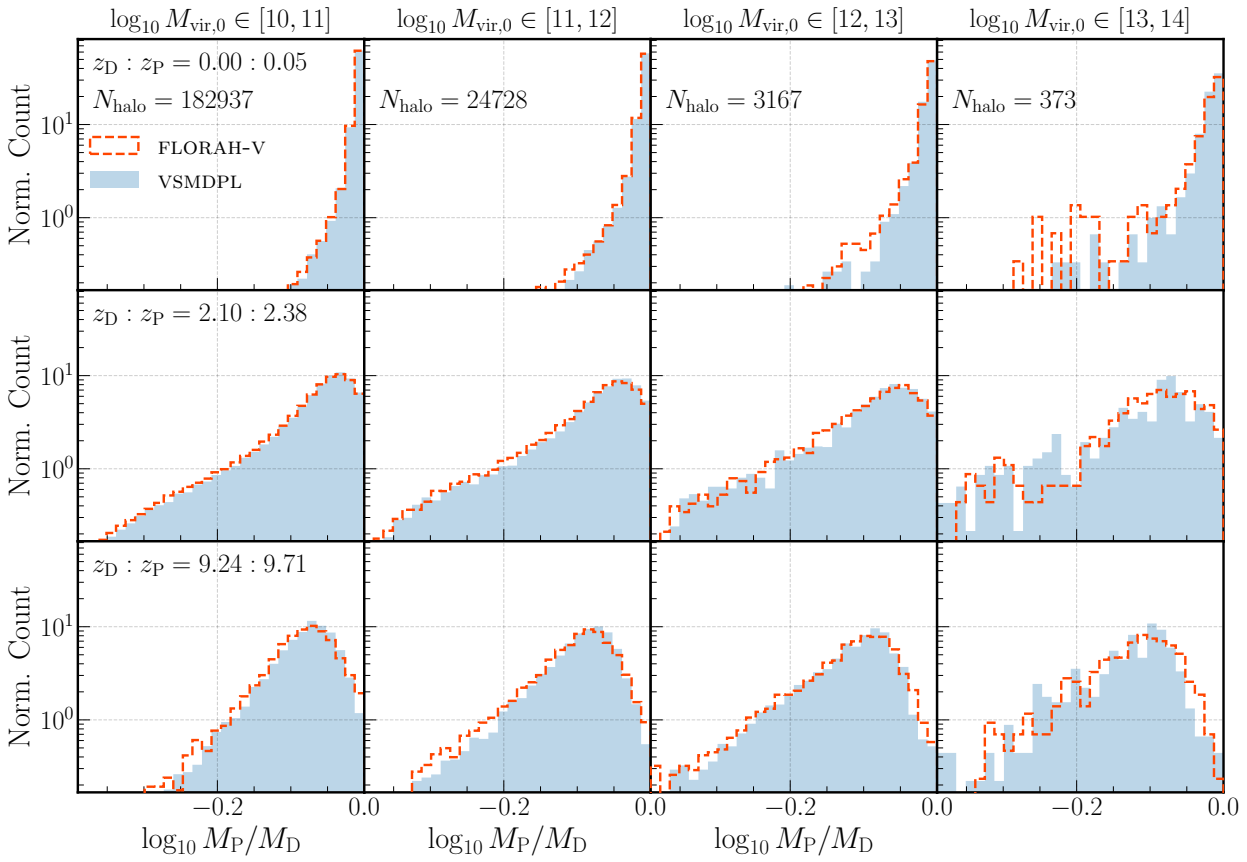

Figure 3.5: The normalized distribution of the progenitor-descendant mass ratio $\log_{10} M_P/M_D$ as modeled by FLORAH-V (blue dashed line) and from the VSMDPL simulation (shaded orange). Each column represents a different root mass bin (in $M_\odot$ unit). Each row represents a different progenitor-descendant redshift slice.

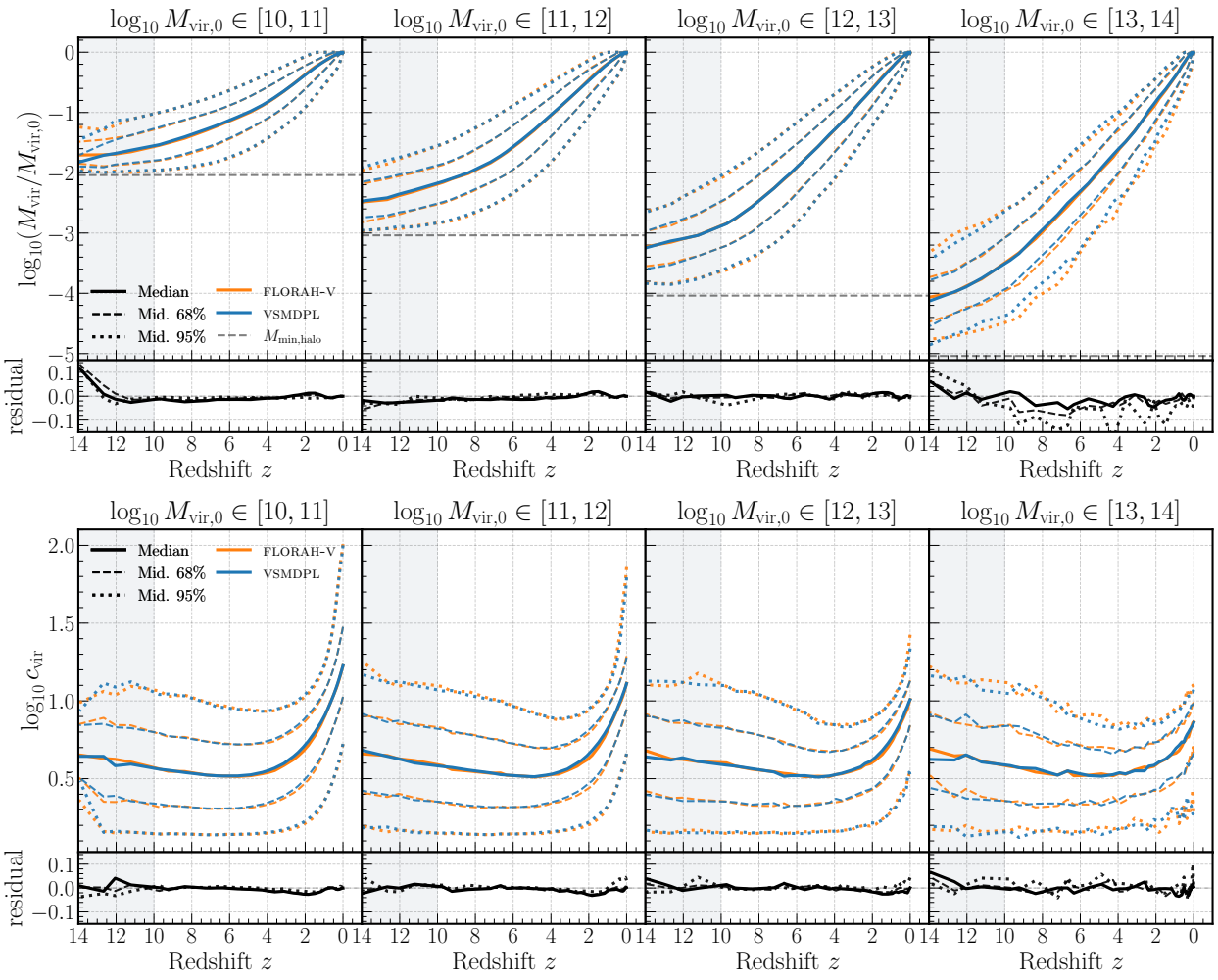

Figure 3.6: The medians, middle 68-% percentile, and middle-95% containment regions of the MAHs (top) and DM concentration histories (bottom) of VSMDPL (dashed blue) and FLORAH (solid orange), along with their residuals (bot), in four mass bins (in $M_\odot$ unit). The residual of the containment region is computed by averaging the residuals of the corresponding upper and lower percentile curves. The shaded gray box ($z > 10$) denotes the "extrapolation region" beyond the maximum training redshift.

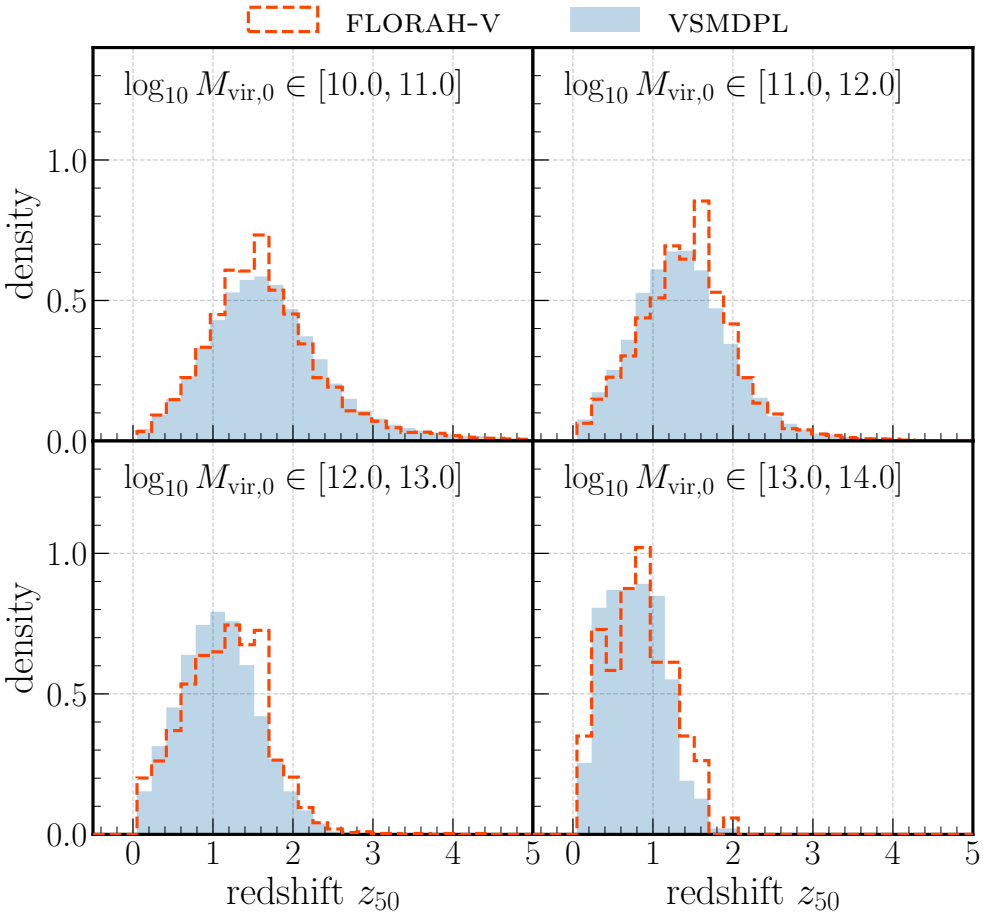

Figure 3.7: The formation redshift z_{50}, defined as the redshift at which the halo forms 50% of its mass, for four mass bins of VSMDPL. Masses are in $M_\odot$ unit.

remove all halos below this mass limit. In all mass bins, the MAHs predicted by FLORAH-V are consistent with those from VSMDPL. As mentioned previously, the last mass bin $\log_{10} M_{\rm vir,0} \in [13.0, 14.0]$ dex contains significantly fewer MPBs, so the containment regions are noisier. In the bottom panels of Figure 3.6, we show the DM concentration $c_{\rm vir}$ histories of VSMDPL and FLORAH. Similarly, the concentration histories predicted by FLORAH agree well with those of VSMDPL. The shaded gray box represents the $z > z_{\rm max,train} = 10$ region, beyond which the MAHs and concentration histories are extrapolated. Except for the first bin, FLORAH-V demonstrates an impressive capability to extrapolate and predict the MAHs and concentration histories up to $z = 14$. In the lowest mass bin, FLORAH accurately extrapolates up to about $z \approx 12$, but it overestimates the MAHs after this point. We attribute this to the fact that this bin is the closest to the resolution limit $M_{\rm min,halo}$, as can be seen from the horizontal dashed line, and caution against extrapolating FLORAH near this limit.

In Figure 3.8, we show the joint distribution between a few chosen redshifts ($z = 0.1, 1, 2, 4, 6, 8$) for $M_{\rm vir}$ and $c_{\rm vir}$ for FLORAH-V-generated (solid orange) and GUREFT-C (dashed blue) MPBs. The 68% and 95% of the $M_{\rm vir}$ joint distributions are shown in the left panel and $c_{\rm vir}$ joint distributions in the right panel. The $M_{\rm vir}$ values at different redshifts are highly and positively correlated. The mass correlation becomes weaker over a wider redshift range, as expected. On the other hand, the $c_{\rm vir}$ values are more correlated at low redshifts (although still much less than $M_{\rm vir}$) and uncorrelated at high redshifts. For both $M_{\rm vir}$ and $c_{\rm vir}$, FLORAH-V can learn the joint distributions and correlations between redshifts present in VSMDPL simulation.

We also emphasize that FLORAH jointly predicts the mass and DM concentration of the halos at each redshift using the conditional normalizing flows. The flows excel at modeling high-dimensional correlations, so additional halo properties (e.g., halo shape, environment density) can naturally be incorporated into FLORAH *without requiring additional assumptions about the data*. In Figure 3.9, we show the median concentration-mass relations at different redshifts, as derived from VSMDPL (dashed lines) and FLORAH-V-generated (solid lines) MPBs. The error bars represent the spread in the concentration-mass relations. The bottom panel shows the residual between the FLORAH-V and VSMDPL median. In general, we see that the concentration-mass relations recovered by FLORAH-V are consistent with VSMDPL across multiple redshifts.

In Figure 3.7, we show the distribution of the formation redshift z_{50}, defined as the redshift at which the halo forms 50% of its mass, for the four chosen mass bins. As mentioned above, low-mass halos tend to form at higher redshift than high-mass halos. We see that the z_{50} distribution by FLORAH-V agrees well with VSMDPL. We emphasize that the formation redshift z_{50} is not explicitly learned by the network but instead derived from the MAHs. This

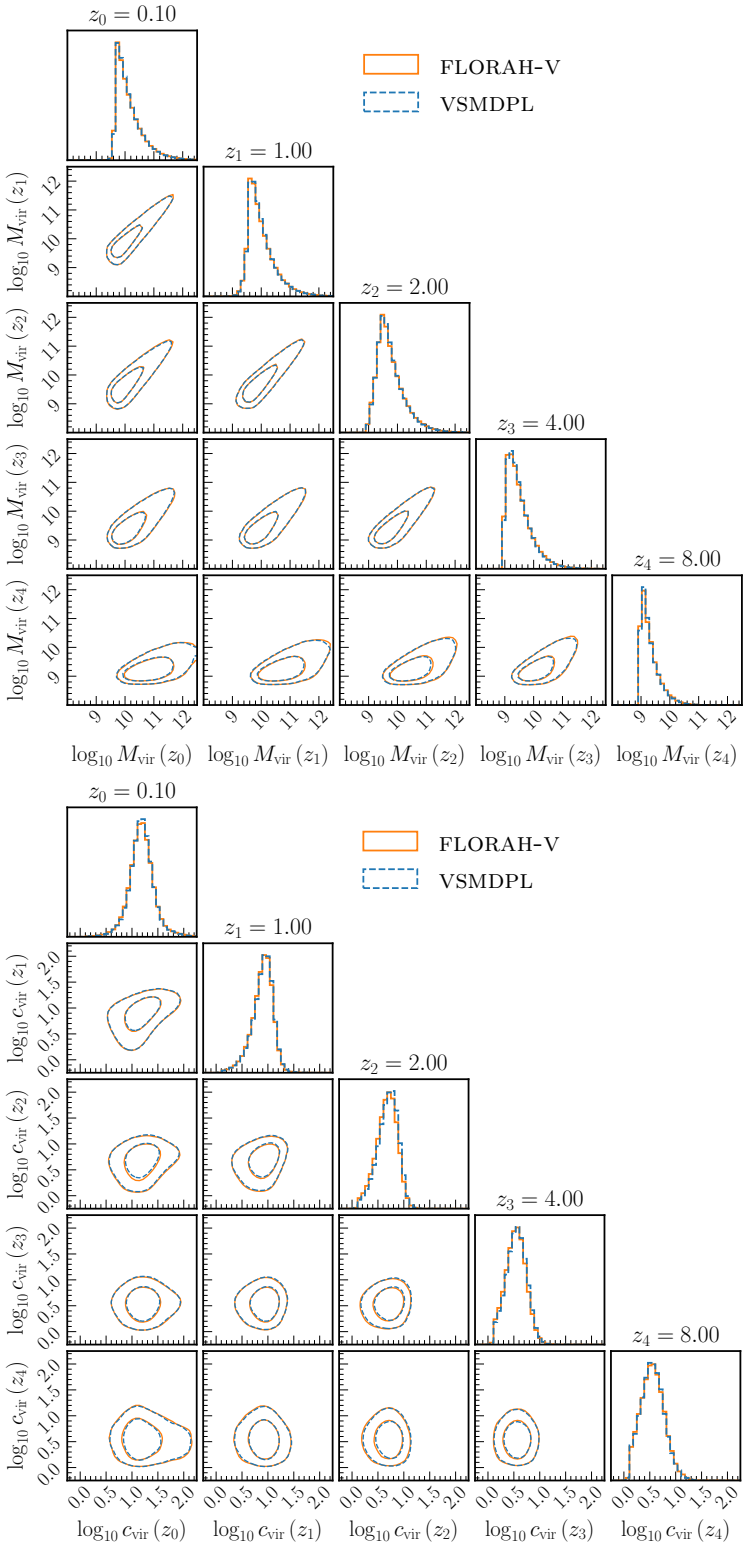

Figure 3.8: The joint distributions of $\log_{10} M_{\rm vir}$ (top) and of $c_{\rm vir}$ (bottom) across a few chosen redshifts ($z = 0.1, 1, 2, 4, 6, 8$) of FLORAH-V (solid orange) and VSMDPL (dashed blue). The contour lines show the 68% and the 95% intervals.

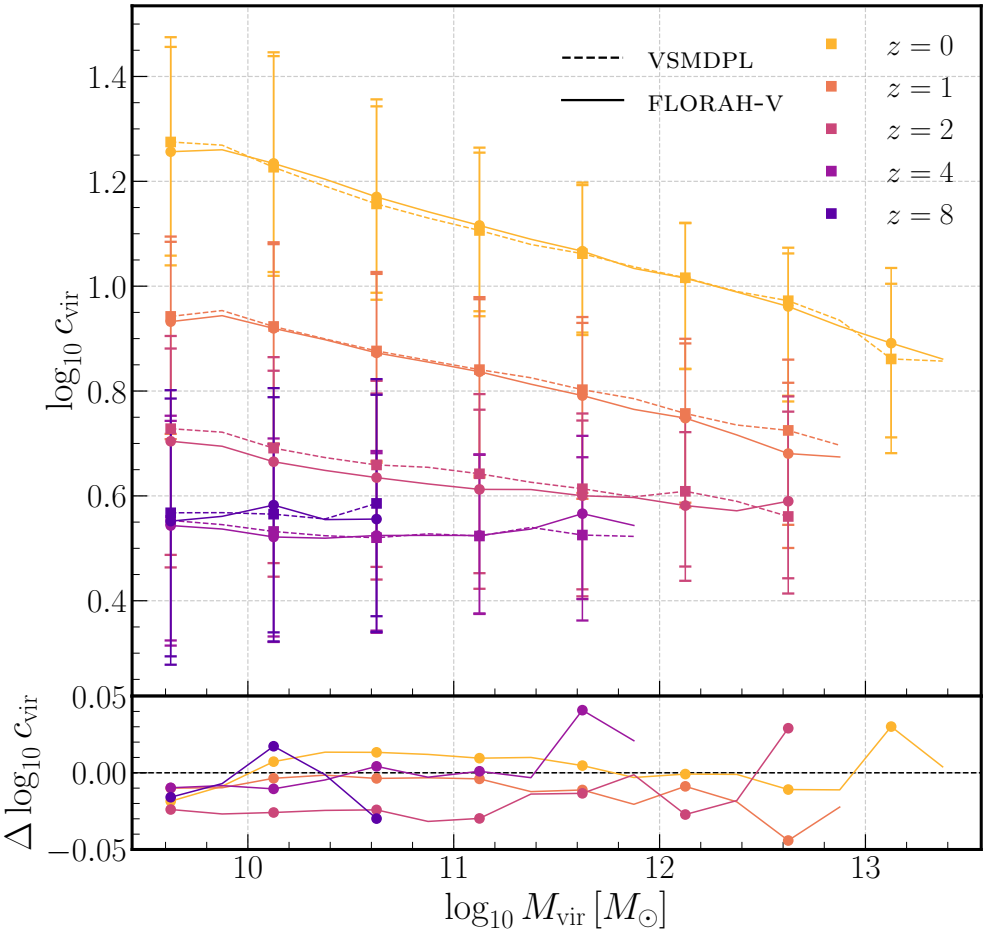

Figure 3.9: The median concentration-mass relations of VSMDPL (dashed lines) and FLORAH-V (solid lines) for five different redshift slices. The error bars denote the spread of the relations, computed from the 16th and 84th percentiles. The bottom panel shows the residual between the FLORAH-V and VSMDPL concentration-mass relations.

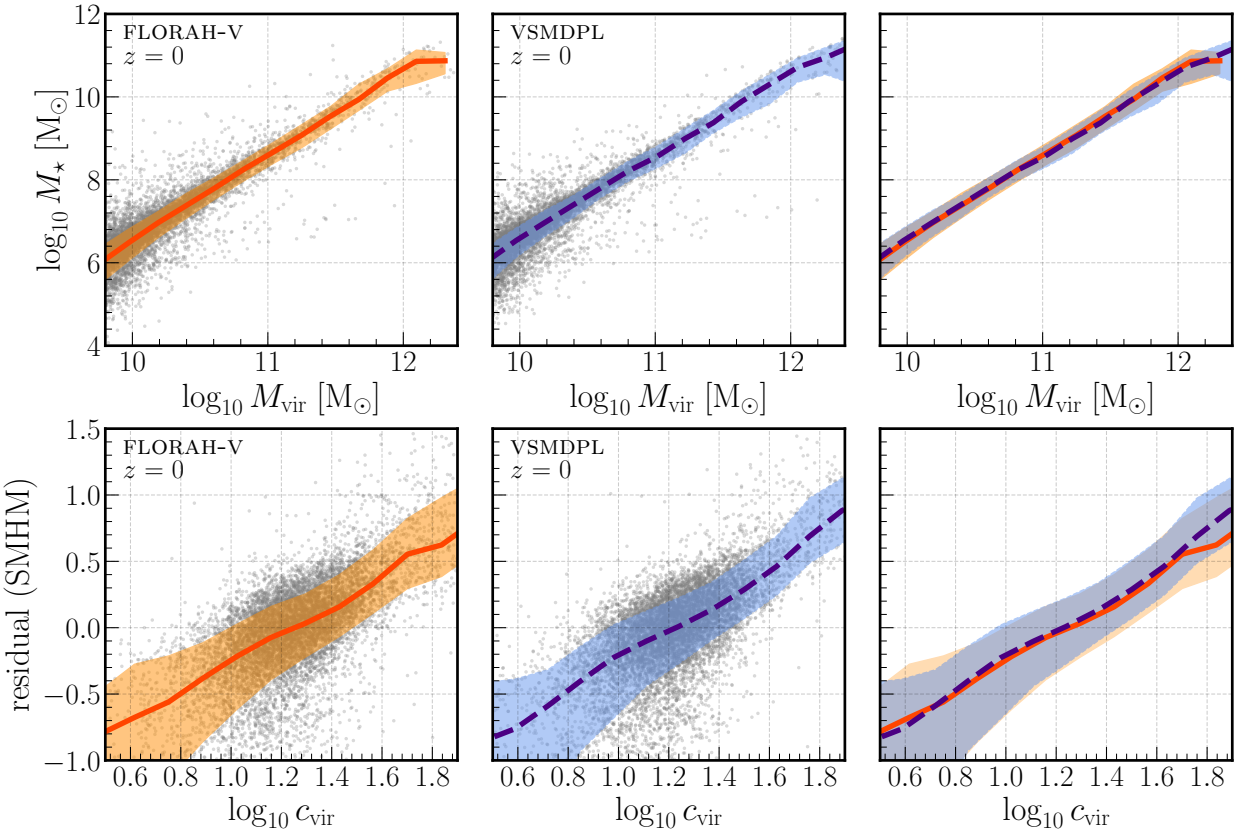

Figure 3.10: Top: The stellar-to-halo mass relation (SHMR) at $z = 0$ computed by the SC-SAM. Bottom: The residual of SHMR, defined as the difference between the $M_\star/M_{\rm vir}$ value for each halo and the median value in its corresponding $M_{\rm vir}$ bin, as a function of the DM halo concentration. In both panels, the shaded regions represent the middle-68 % containment regions.

further shows that FLORAH can capture the "long-range" mass correlation in the MAHs.

Observables and Halo assembly bias

We do not directly observe dark matter halos or merger trees, but only the observable properties of galaxies. In semi-analytic models (SAMs), these properties are predicted by using merger trees as inputs and solving ordinary differential equations for observables like galaxy luminosities and quasi-observables like stellar mass. Thus we need to test that FLORAH-generated assembly histories can be used in place of N-body assembly histories in SAMs to predict these observables. To that end, here we apply the SC-SAM to predict the stellar masses $M_\star$ of galaxies. For a fair comparison with VSMDPL, we input only the MPBs of VSMDPL trees to the SAM. We use both the mass and DM concentration of the generated MPBs. In SC-SAM, the concentration is used in two ways: first, in computing the galaxy

sizes (see [282]), and second, in computing the timescale for disruption of satellite galaxies by tidal forces. Note that since we do not model satellite galaxies in our analysis, the latter point is not directly relevant. We apply the minimum progenitor mass cut $M_{\rm min,halo}$ to both VSMDPL and FLORAH halos. In addition, we do not restrict the redshift range of either VSMDPL and FLORAH halos.

In the top panel of Figure 3.10, we show the stellar-to-halo mass relation (SHMR). Next, we compute the SHMR residual, defined as the difference between the $M_\star/M_{\rm vir}$ value for each halo and its median value in the corresponding $M_{\rm vir}$ bin. The correlation between the SHMR residual and the DM concentration $c_{\rm vir}$ is shown in the bottom panel. Estimating the residual this way removes most of the mass dependence. This demonstrates that the galaxy properties depend on secondary halo characteristics beyond halo mass, which can lead to the phenomenon known as "assembly bias" [269], [271]. In both cases, relations predicted by FLORAH are consistent with VSMDPL. *A SAM run on FLORAH-generated assembly histories accurately reproduces the correlation between stellar mass, halo formation history, and halo concentration.*

However, in the top panel of Figure 3.10, we observe a bifurcation in the SHMR, leading to a secondary population of galaxies with low stellar mass. We theorize that this population represents galaxies which normally grow their mass via mergers. Since we only input MPBs into the SC-SAM, we underestimate the stellar mass of these galaxies. The bifurcation disappears when we input full merger trees of VSMDPL into the SC-SAM, which further supports our theory. Currently FLORAH only generates MPBs, and thus it is unable to capture these merger-driven galaxies. We have also found that using only the MPBs, SC-SAM does not correctly recover the mass of the central supermassive black holes for both VSMDPL and FLORAH. This is expected because the central supermassive black holes grow mass via mergers. We further discuss the limitations of the current framework and future outlook for FLORAH in Section 3.5.

Clustering

To further probe the assembly bias, we compute the galaxy power spectra $P(k)$ of VSMDPL and FLORAH-V-generated galaxy catalogs at $z = 0$. FLORAH does not generate positions, and so for each halo, we take the positions of the corresponding VSMDPL halo in the test dataset at $z = 0$. Thus, only the stellar masses computed by SC-SAM differ between the two catalogs. In the top left panel of Figure 3.11, we show the galaxy power spectra $P(k)$ of VSMDPL (dashed blue) and FLORAH-V (solid orange) catalogs. The ratio between the residual $\Delta P(k) = P_{\rm FLORAH}(k) - P_{\rm VSMDPL}(k)$ and the VSMDPL power spectrum is shown in the bottom panel. We see excellent agreement between the power spectrum of the FLORAH-V

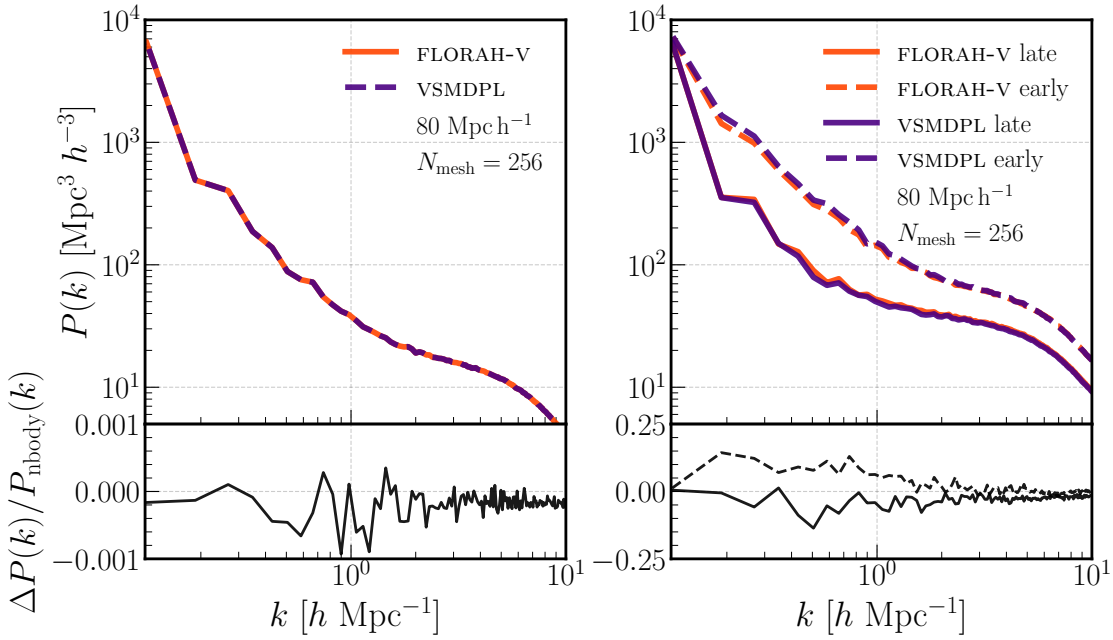

Figure 3.11: Left: The top panel shows the galaxy power spectra $P(k)$ at $z = 0$ of galaxy catalogs from VSMDPL (dashed blue) and generated by FLORAH-V (solid orange). The bottom panel shows the ratio of the residual $\Delta P(k) = P_{\rm FLORAH}(k) - P_{\rm VSMDPL}(k)$ and the VSMDPL power spectrum. Right: The top panel shows the galaxy power spectra $P(k)$ at $z = 0$ of galaxy catalogs from VSMDPL (blue) and generated by FLORAH-V (orange) divided into a late-forming galaxy population (solid) and an early-forming population (dashed). The bottom panel shows the ratio of the residual $\Delta P(k)$ and the VSMDPL power spectrum for the corresponding population. Similarly, the solid line represents the late-forming population, while the dashed line represents the early-forming population.

catalog and that of the VSMDPL catalog. We emphasize that this test is non-trivial because FLORAH is trained on individual halos, independently, and has no notion of any spatial correlations when generating MAHs for a population of clustered halos together.

Studies of assembly bias have also shown that early-forming galaxies exhibit stronger clustering tendencies compared to their late-forming counterparts (e.g., [358]). To investigate this phenomenon, we categorized each galaxy catalog into two groups based on their formation redshift z_{50}: a late-forming population comprised of the first 25% of halos, and an early-forming population consisting of the last 25%. The resulting power spectra $P(k)$ and residuals are displayed in the right panels of Figure 3.11, with the solid line representing the late-forming population and the dashed line representing the early-forming population. The orange lines represent FLORAH-V galaxies, while the blue lines represent VSMDPL galaxies. Our analysis indicates that the power spectra of FLORAH-V galaxies are consistent with those of VSMDPL galaxies for both populations. This further suggests that FLORAH-generated histories can accurately replicate the assembly bias of galaxies.

3.4.2 Learning the combined GUREFT simulations

Using the procedure in Section 3.3.1 to combine the four GUREFT boxes, we extracted 42,774 unique MPBs and applied the data augmentation steps in Section 3.3.1. Note that GUREFT-C contains significantly fewer halos than VSMDPL because the GUREFT boxes have much smaller volumes and particle numbers (as can be seen from Table 3.1). For this reason, we used 38,023 MPBs for the training dataset and 4,751 MPBs for the validation dataset and did not create a test dataset for GUREFT-C. We trained a model, denoted as FLORAH-G, up to a redshift of $z_{\text{max,train}} = 20$. Then we generated 4,751 FLORAH-G MPBs to match the number obtained from GUREFT-C. During the generation process, we applied a similar procedure in Section 3.4.1 to choose the starting halo $x^{(0)} = (\log_{10} M_{\text{vir}}^{(0)}, c_{\text{vir}}^{(0)})$ at $z = 5.89$ and list of scale factors from the *validation* dataset. Similarly, we generate MPBs up to a redshift $z_{\text{max,gen}} = 24 > z_{\text{max,train}}$ to demonstrate the model's extrapolation capability.

We perform additional validation checks similar to Section 3.4.1. In Figure 3.12, we compare distributions of the progenitor-descendant mass ratio $\log_{10} M_D/M_P$ from GUREFT-C (shaded blue) and generated by FLORAH-G (solid orange) for four mass bins (columns) and three redshift slices (rows). We choose the four mass bins such that each bin corresponds to a different GUREFT box, with GUREFT-05 being the least massive bin and GUREFT-90 being the most massive bin. Similar to the VSMDPL example in Section 3.4.1, we observe good agreement between the distributions from GUREFT-C and generated by FLORAH-G.

In Figure 3.13, we present the median, 68%-percentile, and 95%-percentile containment

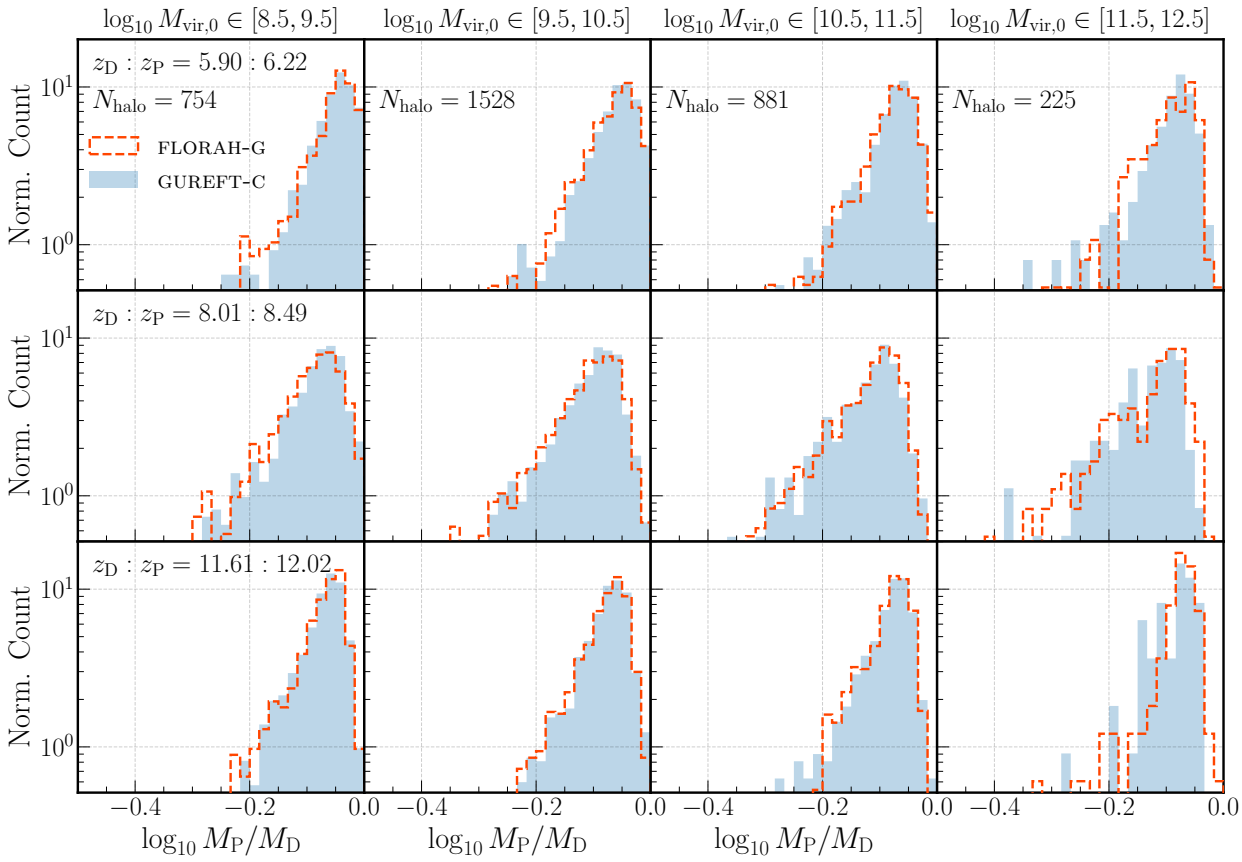

Figure 3.12: The normalized distribution of the progenitor-descendant mass ratio $\log_{10} M_P/M_D$ as modeled by FLORAH-G (blue dashed line) and from the GUREFT-C simulation (shaded orange). Panels are the same with Figure 3.5.

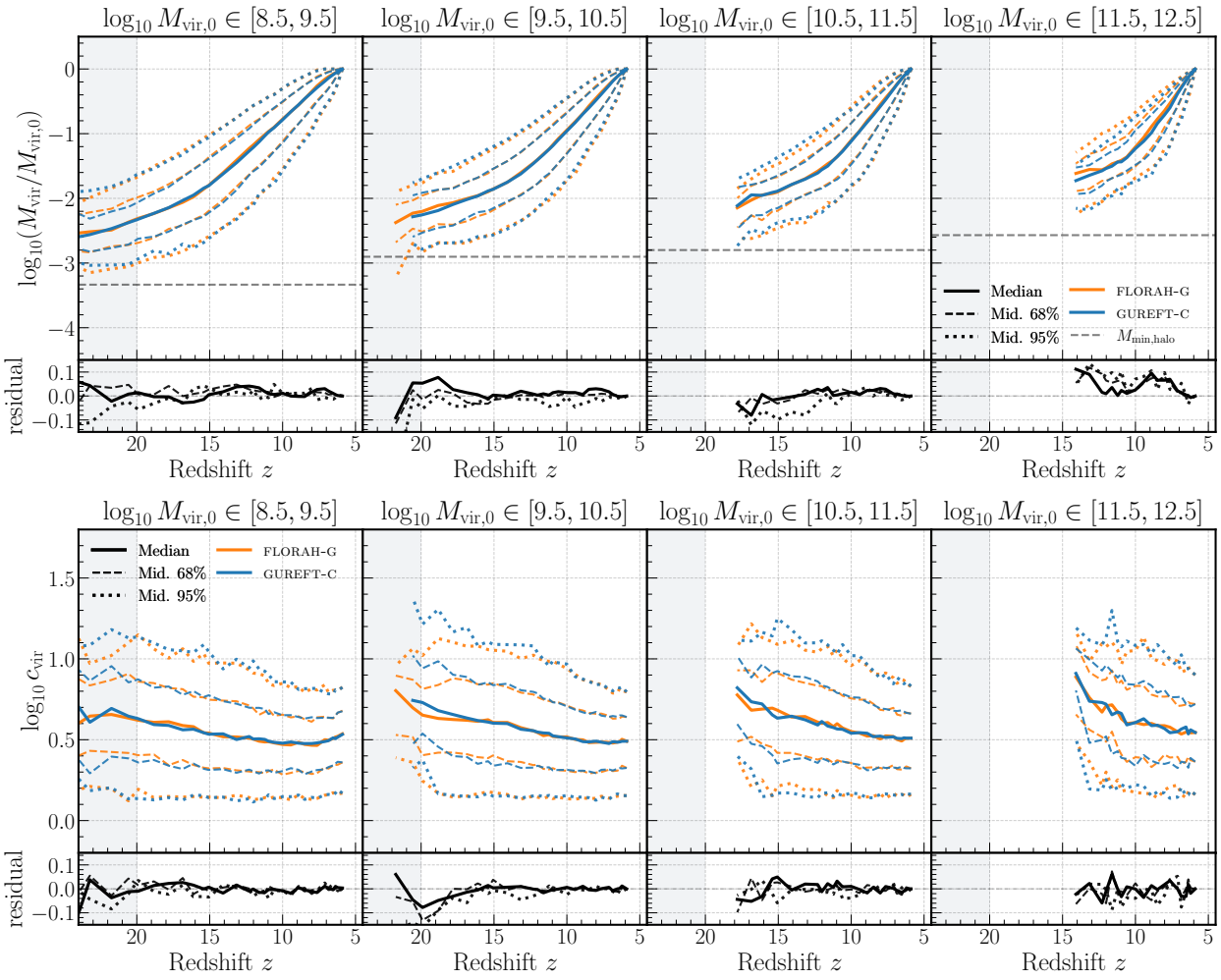

Figure 3.13: The MAHs and DM concentration histories of the MPBs in GUREFT-C and FLORAH-G in four mass bins (in $M_\odot$unit). Panels are the same as Figure 3.6.

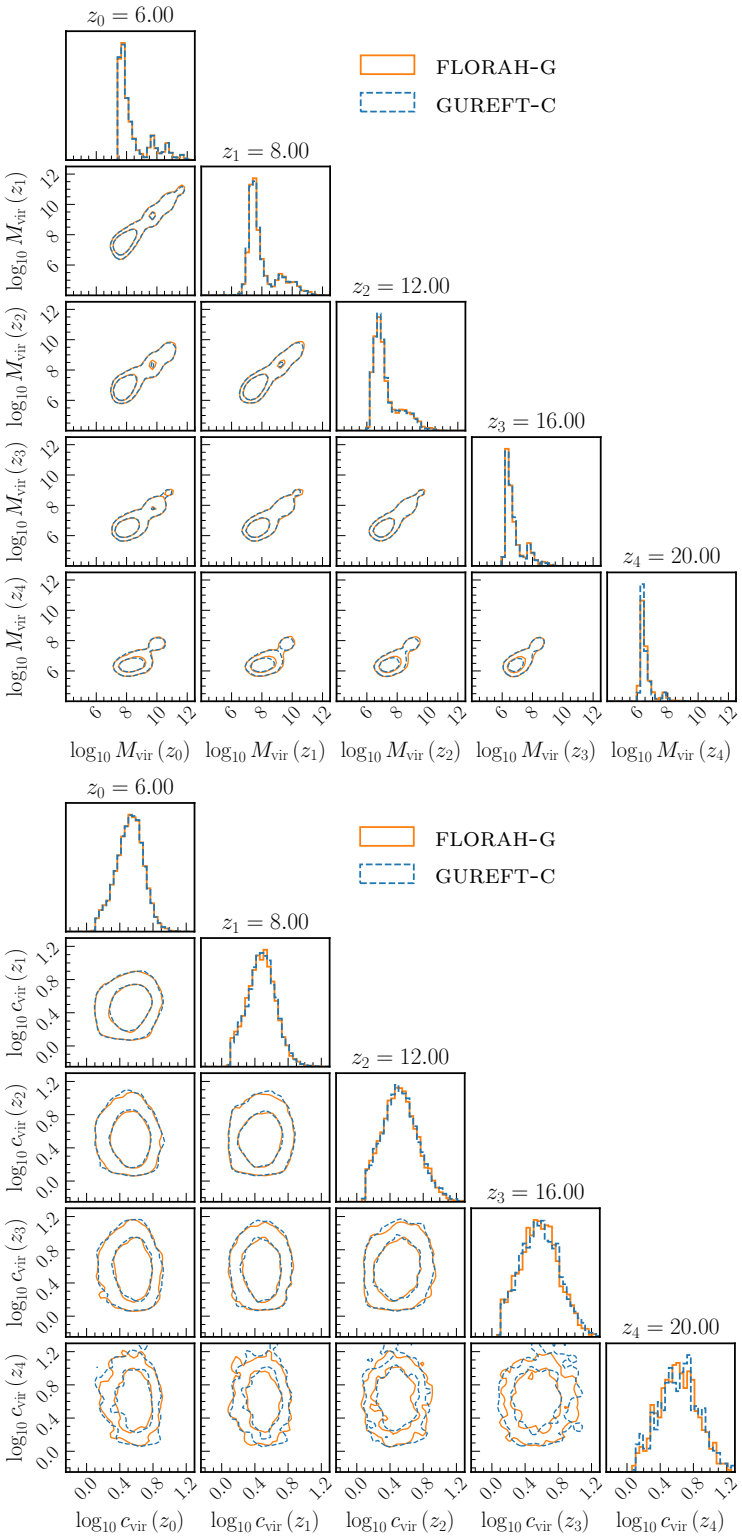

Figure 3.14: The joint distributions of $\log_{10} M_{\rm vir}$ (top) and $c_{\rm vir}$ (bottom) across a few chosen redshifts ($z = 6, 8, 12, 16, 20$) of FLORAH-G (solid orange) and VSMDPL (dashed blue). The contour lines show the 68% and the 95% intervals.

regions of the MAHs (top panels) and the DM concentration histories (bottom panels) for MPBs from GUREFT-C (blue) and generated by FLORAH-G (orange). Similarly to VSMDPL, low-mass halos tend to form at earlier redshifts than high-mass halos. Also as expected, the MAHs of both GUREFT-C and FLORAH-VG plateau out at $M_{\rm min,halo}$. There are a few differences compared to VSMDPL and the low-redshift case. Unlike in VSMDPL, where $M_{\rm min,halo}$ is fixed at 100 times the DM mass of VSMDPL, here $M_{\rm min,halo}$ is determined by the $\log_{10} M_{\rm vir,0}$ at $z = 5.89$ (refer to Table 3.1). choose the same four mass bins as Figure 3.12, with each bin corresponding to a different GUREFT box. Due to a low number of halos in the validation dataset, we only display redshift slices with more than 10 resolved halos, resulting in the termination of the MPBs before $z_{\rm max,train} = 20$ in the last two mass bins. Again, we observe good agreement between the MAHs and concentration histories of FLORAH-G MPBs with those of GUREFT-C MPBs. Moreover, FLORAH-G exhibits a similar capability to extrapolate the MPBs beyond $z_{\rm max,train} = 20$ in the first two mass bins.

In Figure 3.14, we show the joint distributions of $M_{\rm vir}$ (left panels) and $c_{\rm vir}$ (right panels) at selected redshifts ($z = 6, 8, 12, 16, 20$) for FLORAH-G-generated (solid orange) and GUREFT-C (dashed blue) MPBs. The contour lines represent the 68% and 95% intervals. Once again, the distributions of FLORAH-G MPBs match well with those of GUREFT-C MPBs. Note that the mass distribution of GUREFT-C exhibits multiple peaks due to the applied mass cuts in the procedure to combine the four GUREFT-C boxes (see Section 3.3.1).

In Figure 3.15, we show the median concentration-mass relations from the combined GUREFT-C boxes and FLORAH-G-generated MPBs. Similar to the VSMDPL example in Figure 3.9, the error bars denote the spread of the relations, computed from the 16th and 84th percentiles. The bottom panel shows the residual between FLORAH-G and GUREFT-C median relations. Here, we also see that the relations between GUREFT-C and FLORAH-G agree well.

Finally, we show the distribution of the formation redshift z_{50} for the four chosen mass bins of GUREFT-C in Figure 3.16. We see that similarly as in the VSMDPL case, we recover the z_{50} distribution of the GUREFT-C simulations. Again, we emphasize that the formation redshift z_{50} is a derived quantity from the MAHs and not learned directly by the network, indicating that FLORAH can capture the "long-range" mass correlation in the MAHs. We opt not to apply the same SAM analysis shown in Section 3.4.1 and 3.4.1 on FLORAH-G and GUREFT-C MPBs. Given that FLORAH-G can recover the environment-dependent quantities such as the z_{50}, we expect the SHMR, SHMR residual-concentration relation, and galaxy clustering predicted by FLORAH-G to also be consistent with the simulations.

To conclude this section, we developed a procedure to combine multiple N-body simulations with varying resolutions into the training dataset and showed that FLORAH can accurately learn the MAHs and concentration histories of these simulations. The GUREFT

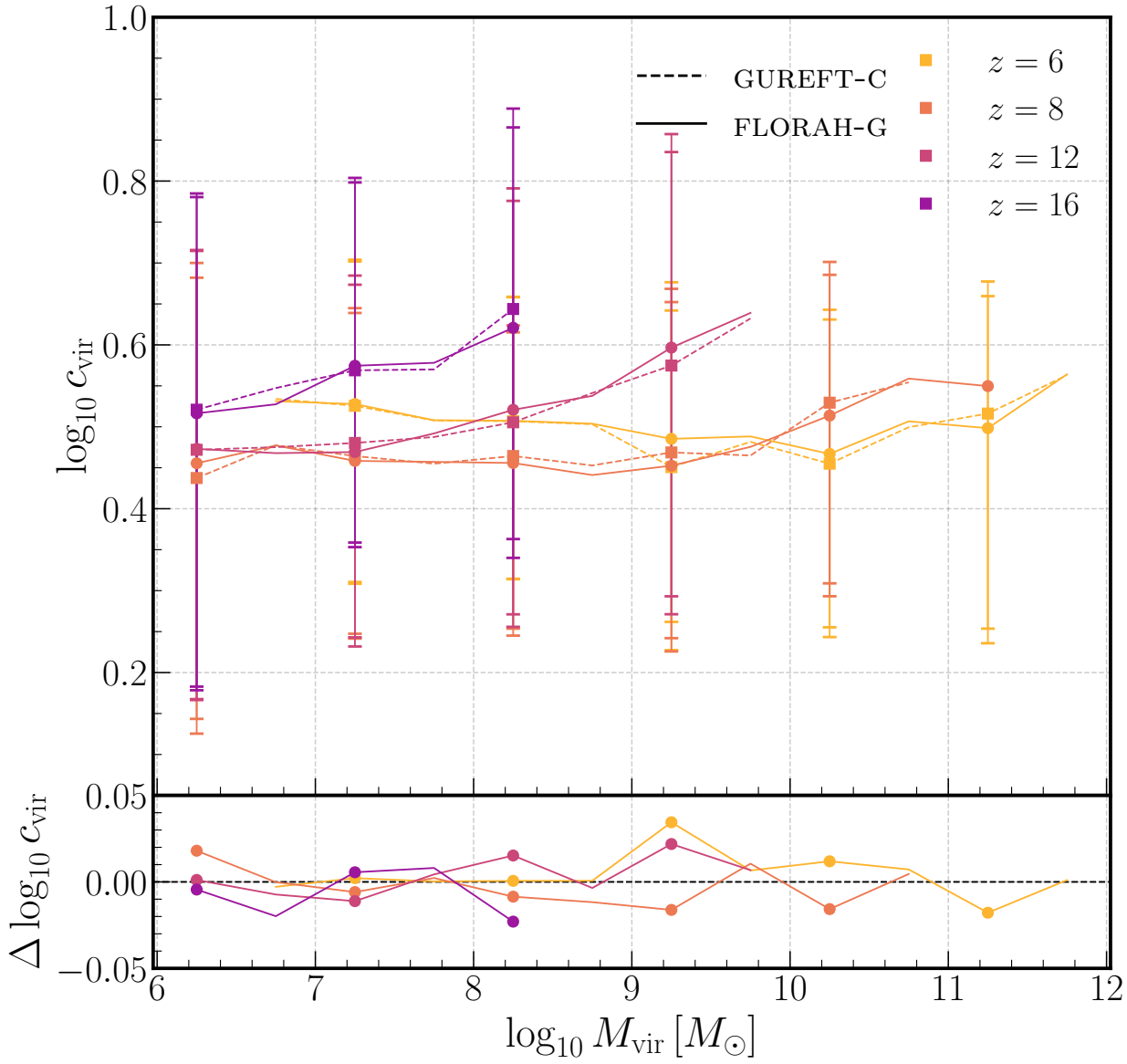

Figure 3.15: The median concentration-mass relations derived from GUREFT-C (dashed lines) and FLORAH-G-generated MPBs (solid lines). Panels are the same with Figure 3.9

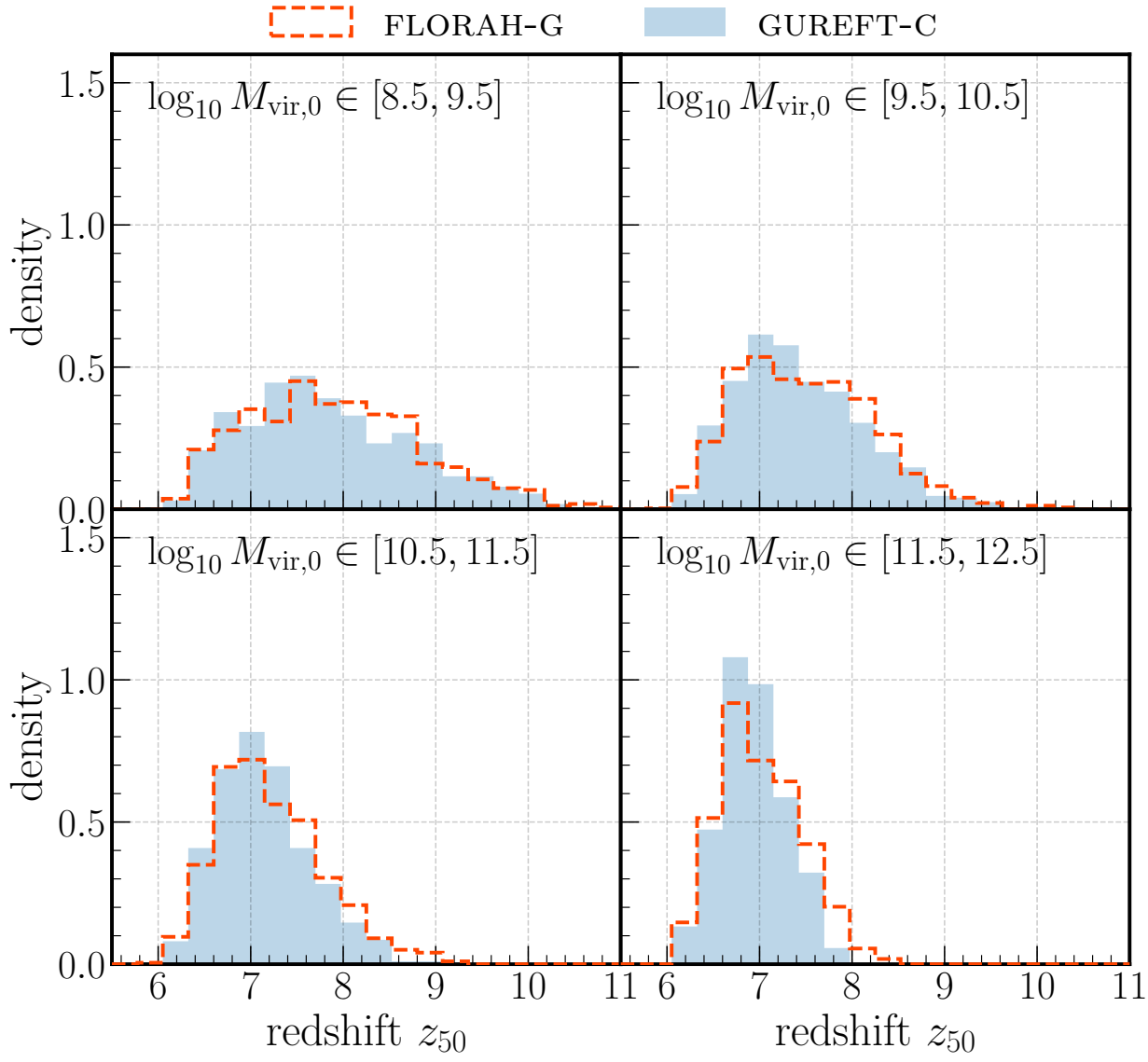

Figure 3.16: The formation redshift z_{50}, defined as the redshift at which the halo forms 50% of its mass, for four mass bins of GUREFT-C. Masses are in $M_\odot$ unit.

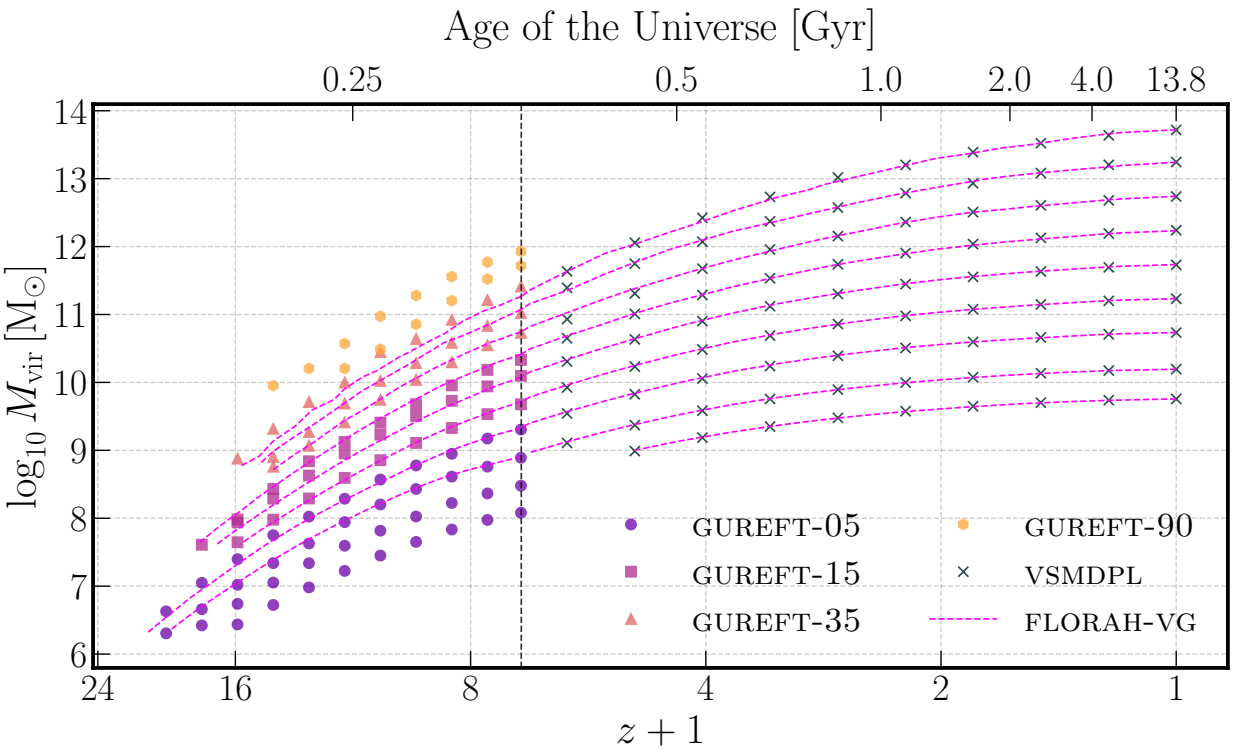

Figure 3.17: The median MAHs of the MPBs of GUREFT, VSMDPL, and FLORAH-VG down to $M_{\rm min,halo} = 100\,M_{\rm DM}$, where $M_{\rm DM}$ is the mass of the DM particle in the corresponding simulations. The dotted vertical line denotes the transition point from VSMDPL to GUREFT-C at $z = 5.89$.

simulations in this example are designed to capture the assembly histories of DM halos at the ultra-high redshifts at an unprecedented temporal resolution. By successfully incorporating the GUREFT simulations into our training dataset, FLORAH becomes an essential tool for understanding and analyzing the formation and evolution of cosmic structures during the early Universe.

3.4.3 Combining GUREFT and VSMDPL

High computational costs make it infeasible to run N-body simulations at the mass resolution required to simultaneously capture the merger dynamics from dwarf galaxies ($10^5 - 10^{10}$ M$_\odot$) to galaxy clusters ($10^{14} - 10^{15}$ M$_\odot$) up to high redshifts and in large volumes. The GUREFT simulations are simulated up to $z \approx 6$ and designed to be complementary to past large volume simulations such as VSMDPL. In this section, we combine the previously trained model, FLORAH-V on VSMDPL and FLORAH-G on GUREFT-C, in an attempt to generate MPBs spanning $z = 0$ to $z \approx 24$.

The generation of the combined assembly histories is done in two steps. First, we applied a similar procedure in Section 3.4.1 to choose the starting halo $x^{(0)} = (\log_{10} M_{\rm vir}^{(0)}, c_{\rm vir}^{(0)})$ at $z = 0$ from the test dataset of VSMDPL. As before, we chose the list of VSMDPL scale factors by sampling every 2 – 6 VSMDPL snapshots from $z = 0$ and used FLORAH-V to generate the MPBs. Here, we only generate the MPBs up to a redshift of $z = 5.89$ and down to a progenitor mass of $M_{\rm min,halo} = 5.89 \times 10^8\,{\rm M}_\odot$, i.e. 100 times the DM mass of VSMDPL. In the second step, for the MPBs that are not terminated before $z = 5.89$, we used FLORAH-G to further generate their progenitors. Note that currently, we do not carry the hidden states between models (i.e. from FLORAH-V to FLORAH-G). This is because the two models are trained independently due to the lack of N-body simulations spanning the entire redshift range from $z = 0$ to $z \approx 24$, and hence their hidden states are not necessarily correlated. Thus here, for the sake of simplicity, we initialize the hidden state for FLORAH-G with zeros but plan on investigating better initialization procedures in the future. We chose a list of GUREFT scale factors by sampling every 2 – 6 GUREFT snapshots from $z = 5.89$ up to $z_{\rm max,gen} = 24$. Again, we set $z_{\rm max,gen} = 24 > z_{\rm max,train}$ of GUREFT-C to demonstrate the model's extrapolation capability. Unlike in VSMDPL, the $M_{\rm min,halo}$ is set based on the progenitor mass at $z = 5.89$. After we generate the full MPB, we apply the post-processing steps in Section 3.3.3. We generated 387,031 MPBs, the same number of trees as the VSMDPL test dataset. For convenience, we denote these combined MPBs as FLORAH-VG MPBs.

We first present the median MAHs of VSMDPL, GUREFT-C, and FLORAH-VG from $z = 0 - 24$ in Figure 3.17. The median MAHs of VSMDPL and GUREFT-C are represented by the data points, while the median MAHs of FLORAH-VG are represented by the dashed lines. The vertical dashed line at $z \approx 6$ represents the transition redshift at which we switch from FLORAH-V to FLORAH-G during the generation procedure. Overall, the FLORAH-VG MAHs lines up nicely with both VSMDPL at low redshifts and GUREFT-C at high redshifts. However, as mentioned in Section 3.2, due to the resolution limit of VSMDPL, less massive root halos of VSMDPL ($\lesssim 10^{10}{\rm M}_\odot$) do not have MAHs that overlap with GUREFT-C. We cannot extend the MPBs of these halos with GUREFT-C and FLORAH-G unless we allow the generation of poorly resolved halos (i.e. halos with fewer than 100 DM particles) during the first step of the generation procedure. Thus, for halos with $\lesssim 10^{10}{\rm M}_\odot$, which corresponds to the host halos of dwarf galaxies, we are unable to reliably generate their assembly histories beyond $z \approx 4 - 5$. The longest FLORAH-VG MPBs have root masses from $1.8 \times 10^{10} - 10^{11}\,{\rm M}_\odot$ (corresponding to the bright end of the dwarf galaxy population) extend to $z \approx 20 - 22$. Finally, it is worth noting that GUREFT-90 and VSMDPL do not overlap due to the mass cut applied when combining the GUREFT boxes (as described in Section 3.3.1). As a result, we do not fully utilize GUREFT-90 during the generation. In future work, we can connect

the massive end of GUREFT-90 with VSMDPL by including an additional simulation in the training dataset.

Next, we examine the generated MPBs in more detail. Since we do not have N-body MPBs that span from $z = 0$ to $z \approx 24$, we divide each FLORAH-VG MPB into a low-redshift component which spans $z = 0 - 10$ and a high-redshift component which spans $z = 5.89 - 25$. We will compare the low-redshift component with VSMDPL and the high-redshift component with GUREFT.

Mass and DM concentration histories at low redshifts

In Figure 3.18, we show the median, 68%-percentile, and 95%-percentile containment regions of the MAHs (top panels) and DM concentration (bottom panels) histories for MPBs from VSMDPL (dashed blue) and FLORAH-VG (solid orange). The MAHs and DM concentration histories of VSMDPL are the same as Figure 3.6. Unlike the VSMDPL MAHs and the FLORAH-G MAHs in Figure 3.6, the FLORAH-VG MAHs do not plateau out at the resolution limit of VSMDPL (horizontal dashed line). This is to be expected because the progenitors of FLORAH-VG halos at high redshifts are generated by the FLORAH-G model and thus have a higher mass resolution. We will show in Section 3.4.3 below that the MAHs at high redshifts are indeed consistent with GUREFT-C. However, we note some unphysical features in the MAHs and concentration histories. The MAHs of VSMDPL and GUREFT-C do not line up perfectly at the transition redshift $z = 5.89$, creating "kinks" in the 95% of the MAHs in the first two mass bins. This is particularly clear in the bottom panels, where the kinks can also be seen in the third mass bins and the median and 68% of the DM concentration histories. We defer the discussions of the limitations of our approach in more detail in Section 3.5, after we compare the high-redshift components of FLORAH-VG to GUREFT-C.

Mass and DM concentration histories at high redshifts

In Figure 3.19, we compare the high-redshift components of the FLORAH-VG MPBs with GUREFT-C MPBs. As above, the median, 68%-percentile, and 95%-percentile containment regions of the MAHs (top panels) and DM concentration (bottom panels) histories. Because GUREFT-C contains significantly fewer MPBs than generated by FLORAH-VG (since FLORAH-VG MPBs are generated from the root halos of VSMDPL), we subsample FLORAH-VG MPBs to match the number of MPBs in GUREFT-C for a fair comparison. Similarly to Figure 3.13, we do not include redshift slices with fewer than 10 resolved halos, resulting in the MAHs and concentration histories in the last two bins terminating before $z_{\text{max,train}} = 20$. In general, both the MAHs and concentration histories of FLORAH-VG agree well with GUREFT-C. Unlike in

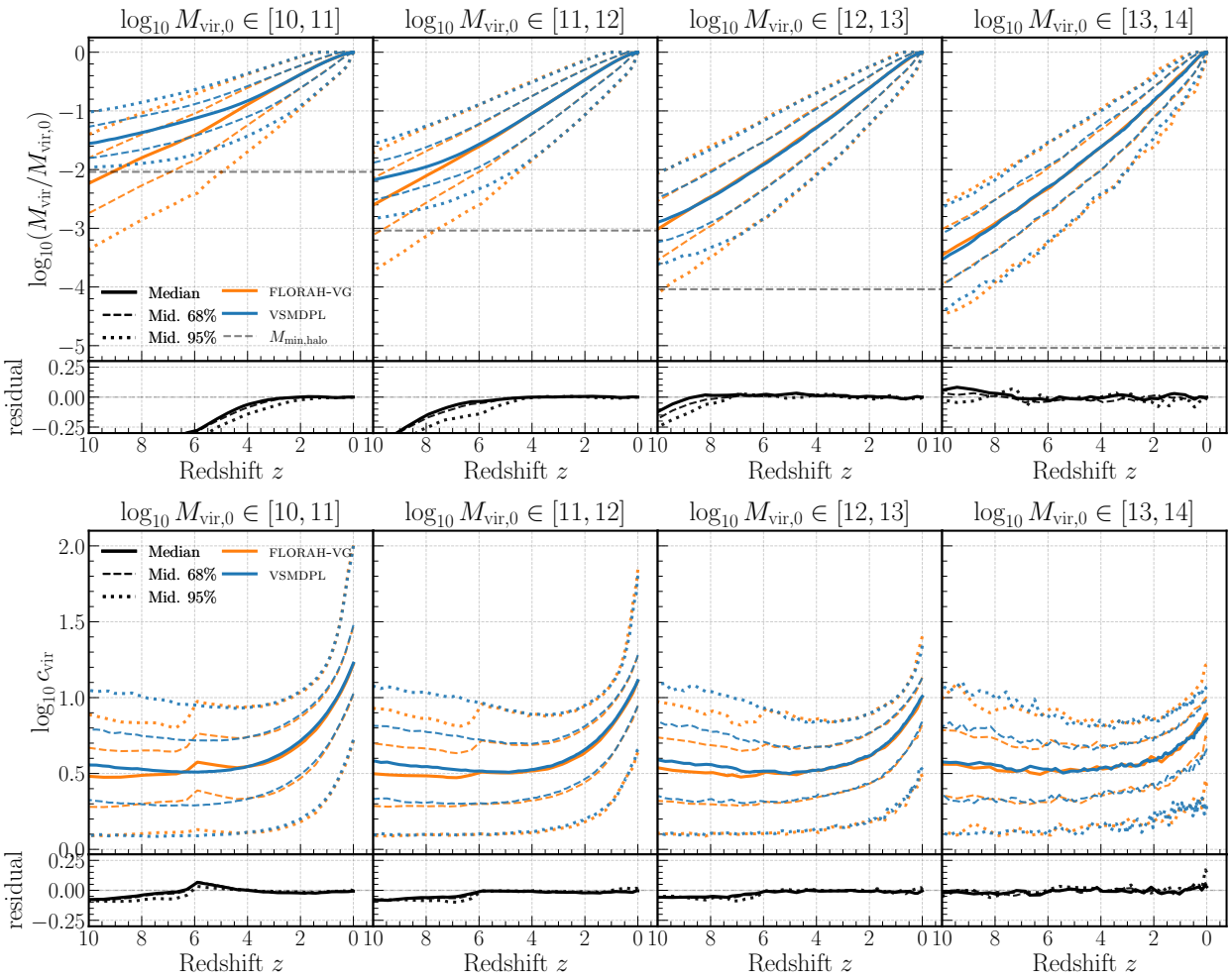

Figure 3.18: The MAHs and DM concentration histories of the MPBs in VSMDPL and FLORAH-VG in four mass bins (in $M_\odot$unit. Panels are the same as Figure 3.6. The disagreement between the VSMDPL MAHs and FLORAH-VG MAHs at high redshift is expected. Similarly to Figure 3.6, the VSMDPL MAHs plateau out at the resolution limit of VSMDPL. On the other hand, at $z \gtrsim 6$, FLORAH-VG MAHs do not plateau out since they are generated by FLORAH-G, which is trained on GUREFT-C, and thus have higher resolutions than VSMDPL.

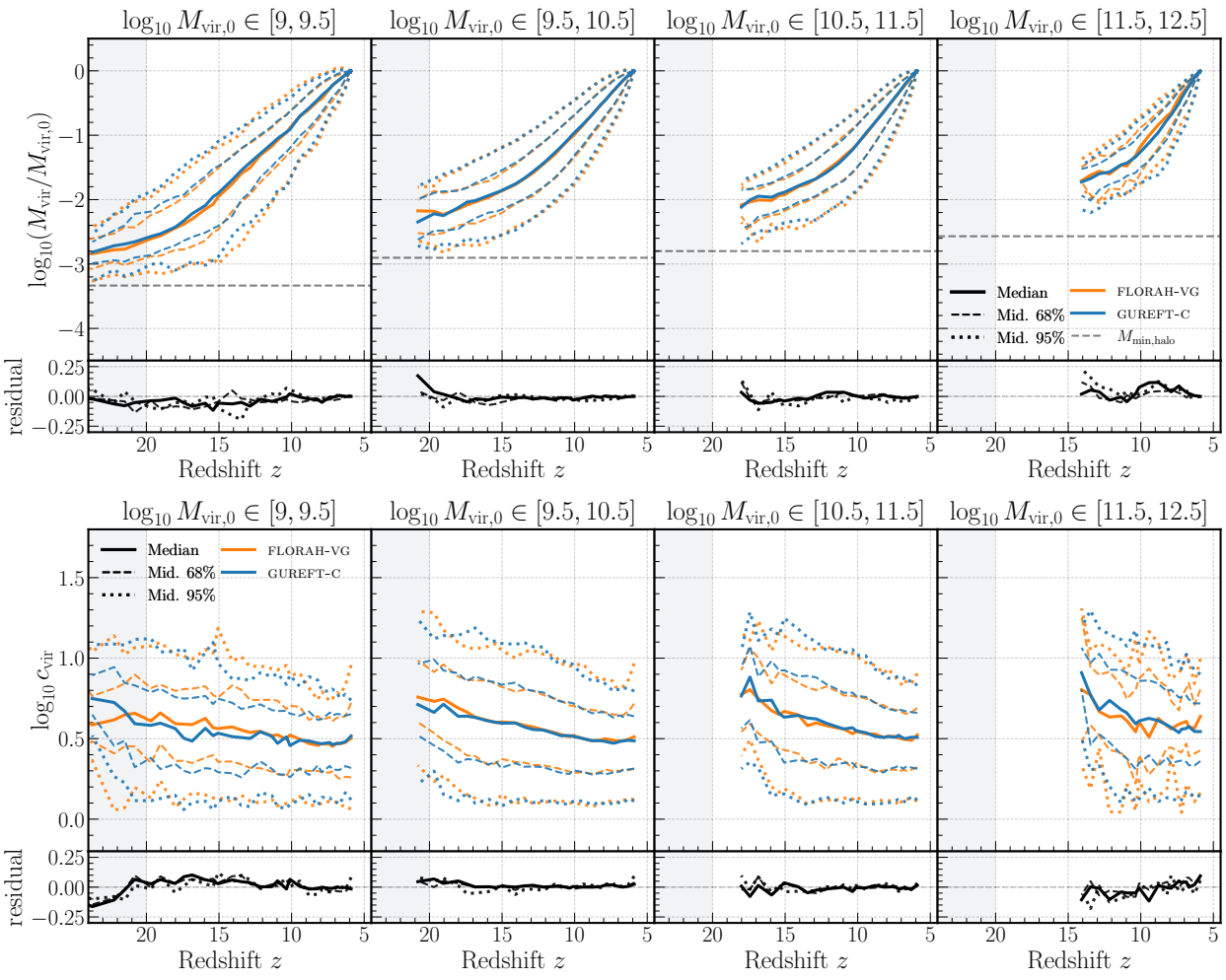

Figure 3.19: The MAHs and DM concentration histories of the MPBs in GUREFT-C and FLORAH-VG in four mass bins (in $M_\odot$unit. Panels are the same as Figure 3.6.

the low-redshift case, the MAHs of both GUREFT-C and FLORAH-VG plateau out at $M_{\rm min,halo}$ as expected. As a reminder, here $M_{\rm min,halo}$ is determined by the $\log_{10} M_{\rm vir,0}$ at $z = 5.89$ (refer Table 3.1). Note that we still see the effect of the sharp transition between VSMDPL and GUREFT-C at $z = 5.89$ in the concentration histories, though to a much lesser extent.

3.5 Discussion

3.5.1 Comparison with previous works

Previous approaches to modeling the cosmological MAHs of DM halos typically fit a parameterized functional form over the entire simulated MAH population (e.g. [334], [349], [363], etc.). For example, [363] assumes the MAHs take the form:

$$M(z) = M_0(1+z)^{\beta} e^{-\gamma z}, \tag{3.7}$$

where $M_0 = M(0)$ is the mass of the root halo, and $\theta \equiv (\beta, \gamma)$ are free parameters of the fit. One can also imagine fitting the evolution of the DM concentration $c_{\rm vir}(z)$ or other properties of the halos in a similar manner. As shown in the aforementioned works, these approaches can also provide a reasonably accurate representation of the MAH and capture population-level statistics such as the mean MAH for a given root halo mass. To capture the full distribution (i.e., the scatter) of the MAH population, one can model the full distribution of the fit parameters $P(\theta)$, as shown in [349]. The dependency of θ on the root halo properties such as M_0 and c_0 can be included as a conditional probability, i.e. $P(\theta|M_0, c_0)$.

The FLORAH framework can be considered as a non-parametric modeling method for the MAH, with a few notable advantages. First, the parametric modeling approaches above assume a functional form not only for the mean of the MAH but also the scatter (and other quantiles) of the MAH population distribution, as this stochasticity is folded into the sampling of the fit parameters $P(\theta|M_0, c_0)$ at $z = 0$. It is unclear if these all should follow the same distribution. Thus, non-parametric methods such as FLORAH can provide more flexibility in modeling the MAH distribution. Moreover, parametric methods may require additional assumptions to incorporate more than one halo property. For example, to fit both the MAH and the DM concentration histories, one needs to assume how the mass and concentration are correlated, e.g. via the halo concentration-mass relations. On the other hand, FLORAH can jointly fit both the mass and DM concentration *without requiring additional assumptions about the data*. Indeed, we show in Figures 3.9 and 3.15 that MPBs generated by FLORAH follow the same halo concentration-mass relations as the simulations. Additionally,

incorporating additional halo properties (e.g., the halo shape or environment information) into FLORAH is straightforward, since the conditional normalizing flows can easily model high-dimensional distributions. We plan to explore different sets of halo properties in future work.

As shown in Section 3.4, FLORAH can correctly capture the progenitor-descendant mass ratios at each time step and model the MAH at the individual halo level. The distribution of the mass (and DM concentration) at each redshift z is modeled independently with the conditional normalizing flows, thus allowing for "jumps" in the MAH to occur, representing DM halo mergers, which can play an important role in galaxy formation models. In contrast, the parametric modeling approaches assume the MAH to be a smooth function, so there is no stochasticity between redshifts once the fit parameters θ are sampled from $P(\theta|M_0, c_0)$ at $z = 0$. Though it is possible to increase the fidelity of the parametric fits by resampling θ at different z, this requires modeling $P(\theta|M_0, c_0, z)$, which will be much more difficult. Halo-level properties such as the progenitor-descendant mass ratios are important for certain modeling applications of black hole seeding in SC-SAM, which we will explore with FLORAH in future works. We note that EPS-based approaches can also model the stochasticity between redshifts in a similar autoregressive manner, by sampling the conditional mass function from the EPS formalism. However, in EPS-based approaches, the MAH is usually modeled as a Markov Chain, in which the mass of the progenitor is modeled using only the mass of its immediate descendant. FLORAH extends this so that the progenitor mass is modeled using the entire assembly history.

3.5.2 Current limitations and future outlook

We showed in Section 3.4.1 that when trained on a single N-body simulation, FLORAH can accurately recover the MAHs and concentration histories. This applies similarly when multiple N-body simulations with the same redshift range but varying mass resolutions are combined into a single training dataset (Section 3.4.2). We further showed that FLORAH-generated MAHs can be input into the SC-SAM to predict observable properties and assembly bias of galaxies. However, because the current framework only generates the MPBs of merger trees, FLORAH cannot yet capture correctly merger-driven properties such as the stellar mass of some population of galaxies or the mass of supermassive black holes. To capture these properties, we must include secondary branches as well as other sub-branches in the generation process. In ongoing work, we are extending it to generate full merger trees by, for example, modeling the probabilities of branching events and utilizing graph generative models.

In Section 3.4.3, we combined the FLORAH-V and FLORAH-G model to generate assembly

histories from $z = 0$ to $z \approx 24$. Overall, we found that the generated MAHs and concentration histories agree with the N-body simulations. However, the current approach has two main drawbacks.

First, due to the lack of N-body simulations spanning the redshift range from $z = 0$ to $z \approx 24$, FLORAH-V and FLORAH-G are trained independently. Thus, their hidden states are not necessarily correlated and cannot be carried from one model to the other (i.e. from FLORAH-V to FLORAH-G). As a result, the information carried by the hidden states of FLORAH-V is lost during the transition. We plan on investigating better initialization procedures in future work. For example, instead of transitioning at $z = 5.89$, we may better utilize the overlapping redshift range between VSMDPL and GUREFT. We found that the current range, $z > 5.89$, to be insufficient because it includes too many poorly resolved halos of VSMDPL (i.e. those with fewer than 100 DM particles), which create an inconsistency between the MAHs of VSMDPL and of GUREFT.

Second, the current approach for combining the simulations can result in some unphysical features in the MAHs and concentration histories at the transition redshift $z = 5.89$ between the two simulations. This is especially clear in the concentration histories, where we see that the concentration histories of VSMDPL are consistently higher than FLORAH-VG at $z > 5.89$. However, from Figure 3.19, it is clear that the concentration histories of FLORAH-VG is consistent with GUREFT-C. This implies that there is a systematic difference between the concentration histories of GUREFT-C and VSMDPL. We attribute this difference to the absence of unresolved, low-mass progenitor halos in VSMDPL at these redshifts. At high redshifts, $c_{\rm vir}$ tends to be lower for lower-mass halos, so missing the low-mass halo population will artificially push the distribution of $c_{\rm vir}$ towards the high values. This is supported by the fact that the lower-68% and lower-95% of the concentration histories are less affected compared to the upper-68% and upper-95%. Although a more sophisticated generation procedure may mitigate the sharp transitions, we attribute the problem mainly to the inconsistency in the MAHs and concentration histories of the two training simulations. We refer to [154] for a more detailed diagnosis of this inconsistency and comparison between GUREFT and the MultiDark simulations (including VSMDPL).

Both of the above drawbacks can be mitigated by improving the simulations. In future work, we are considering the possibility of further running GUREFT to a redshift of 4 and with a larger box size (e.g. 120 $\mathrm{Mpc\,h^{-1}}$). This will extend the overlapping redshift range between the two simulations and ensure that VSMDPL has a sufficient number of resolved halos in this range, thus improving the consistency between the assembly histories. In addition, this enables the assembly histories of halos with less than $10^{10}\,\mathrm{M_\odot}$ to be extended using a GUREFT-based FLORAH model.

3.6 Conclusion

In this chapter, we developed FLORAH, a deep generative model based on recurrent neural networks and normalizing flows, to generate the mass assembly histories (MAHs) of dark matter halo merger trees. We trained two models independently on the VSMDPL and GUREFT N-body simulations and assessed the performance of the models to recover the MAHs, concentration histories, and observable and semi-observable properties derived from semi-analytic models such as stellar mass. Our main findings are summarized as follows:

- In Section 3.4.1, we found that the FLORAH-V model, which is trained on the VSMDPL simulation up to redshift 10, can accurately capture the MAHs and concentration histories across a wide range of host halo masses. Additionally, when operating well above the resolution limit of the simulation, FLORAH-V robustly extrapolates the histories up to redshift 14, beyond the training redshift. We applied the Santa Cruz semi-analytic model to FLORAH-generated histories and showed that FLORAH can recover correlations between galaxy properties and halo assembly history, such as the correlation between stellar-to-mass ratio residual and halo concentration, as well as the galaxy matter power spectrum. FLORAH is the first machine learning framework that accurately learns and incorporates the correlations between halo properties and merger history.

- In Section 3.3.1, we developed a procedure to incorporate multiple simulations with the same redshift range and varying mass resolutions into a single training dataset. In Section 3.4.2, we trained the FLORAH-G model on the GUREFT suite of simulations, comprising four boxes (GUREFT-05, GUREFT-15, GUREFT-35, and GUREFT-90), and successfully demonstrated FLORAH's ability to handle this variation. The GUREFT simulations were purposefully crafted to capture the evolution of dark matter halos at the ultra-high redshift Universe and at an unprecedented temporal resolution. By successfully learning the assembly histories of these simulations, FLORAH proves to be a versatile tool for studying the formation and evolution of cosmic structures during the initial phases of the Universe.

- In Section 3.4.3, we developed a procedure to concatenate multiple FLORAH models. We used FLORAH-V to generate the assembly histories from $z = 0$ to $z = 5.89$ and FLORAH-G to further generate the histories to $z \approx 20 - 24$. We showed that despite the simplicity of our approach, we were able to generate the mass assembly histories from $z = 0$ to an ultra-high redshift $z \approx 24$ for halos with root mass $\lesssim 10^{10}\,\mathrm{M}_\odot$, which corresponds to the bright end of the dwarf galaxy population. We were unable to

reliably generate assembly histories below this mass due to the resolution limitation of VSMDPL. In addition, during the generation process, we found an inconsistency in the dark matter concentration histories of VSMDPL and GUREFT at $z \gtrsim 6$, which may be attributed to the missing of unresolved progenitor halos in VSMDPL (refer to [154] for a more detailed discussion). This results in sharp kinks in the concentration histories at the transition redshift $z = 5.89$ between the two FLORAH models. These limitations can be overcome by improving the overlapping between GUREFT-C and VSMDPL, e.g., by running GUREFT-C to a lower redshift and with a larger volume.) Despite the current limitations, this represents the first step towards generating assembly histories at a mass resolution required to simultaneously capture the merger dynamics from dwarf galaxies to galaxy clusters up to high redshifts and in large volumes. At present, this is far beyond the capability of numerical simulations.

- In Section 3.5, we discuss the advantages of FLORAH over traditional parametric modeling approaches for the MAHs. Like many machine learning methods, FLORAH is non-parametric and does not require assumptions on the functional form of the mass and DM concentration assembly histories. It is also straightforward to include additional halo properties beyond the mass and DM concentration since the normalizing flows employed for density modeling excel at capturing high-dimensional distributions. In addition, FLORAH models the MAHs in an autoregressive manner, sampling the halo mass and concentration at each redshift based on the entire halo history. This allows FLORAH to capture variations between redshifts, such as the progenitor-descendant mass ratios, more easily.

FLORAH represents an exciting and promising initial step towards the development of a machine learning-based framework for generating full merger trees. Such a framework has the potential to revolutionize our understanding of how galaxies form and evolve by allowing for the exploration of different galaxy formation scenarios with excellent computational efficiency at unprecedented accuracy. To achieve this goal, we intend to expand FLORAH's capabilities to generate secondary branches and other sub-branches. In addition, we aim to develop an emulator that can generate merger trees by training FLORAH on simulations featuring different cosmological parameters. Doing so will enable FLORAH to learn and capture the dependence of halo assembly histories on cosmology.

Software and Data Availability

This research made use of the `corner` [364], `IPython` [319], `Jupyter` [320], `Matplotlib` [321], `NumPy` [323], `nflows` [322], `PyTorch` [324], `PyTorch Lightning` [325], `SciPy` [365], and `ytree` [366] software packages.

The code version used for this chapter is available at https://github.com/trivnguyen/florah. The VSMDPL training dataset in this article uses public simulations from the CosmoSim and MultiDark database available at https://www.cosmosim.org/. The GUREFT training dataset will be shared on reasonable request to the corresponding author. The output data generated by the FLORAH models will be shared on reasonable request to the corresponding author.

Chapter 4

Emulating subhalo populations under alternative dark matter scenarios with Diffusion Models

Disclaimer

The content of this chapter is a paper draft that is currently under internal revision of the DREAMS collaboration. The lead authors are Tri Nguyen, Francisco Villaescusa-Navarro, Siddharth Mishra-Sharma, and Carolina Cuesta-Lazaro. Following this revision, the paper will be submitted to the Astrophysical Journal and uploaded to arXiv.

Abstract

To maximize the scientific return of cosmological surveys, it is important to model the relationship between galaxies, subhalos, and their host dark matter halos. Many different techniques have been developed to accomplish this, from Halo Occupation Distribution (HOD) models to hydrodynamic simulations passing via empirical and semi-analytic models. Hydrodynamic simulations make accurate theoretical predictions but are computationally expensive; HODs, on the other hand, are computationally cheap, but the accuracy of their predictions is limited. In this work, we develop a model that can paint galaxies/subhalos on top of dark matter halos, with predictions as accurate as those of hydrodynamic simulations but with computational costs similar to HODs. Our model combines normalizing flows, diffusion models, and transformers to predict subhalo positions, velocities, and internal properties as a function of astrophysics and warm dark matter mass. We train our model

on high-resolution zoom-in hydrodynamic simulations from the DREAMS project and compare summary statistics from the simulations against those derived from samples emulated from our model. We show that our model captures the complex relation between subhalo positions, velocities, and internal properties. Our method can be used for a large variety of downstream tasks, from galaxy clustering to strong lensing studies.

4.1 Introduction

The abundance and spatial distribution of galaxies, and subhalos in general, contain a wealth of information about the composition and laws governing the Universe. For instance, different dark matter (DM) models will predict different subhalo abundances. If we want to maximize the information we can extract from these surveys (through e.g. galaxy clustering or strong lensing studies), we need accurate theoretical predictions for galaxy/subhalo properties as a function of cosmology, astrophysics, and DM properties.

There are different ways to obtain theoretical predictions for the abundance and spatial distribution of galaxies/subhalos. On one end of the spectrum, we have full hydrodynamic simulations that make very precise theoretical predictions about the phase-space distribution of galaxies and their intrinsic properties (e.g. stellar mass, metallicity, neutral hydrogen mass, etc) but that are very computationally demanding. On the other end of the spectrum, we have Halo Occupation Distribution (HOD) models that can *paint* galaxies on top of DM halos from computationally cheaper N-body (i.e, gravity-only) simulations at the expense of limited precision on the spatial distribution of galaxies and with the additional challenge of modeling clustering and intrinsic galaxy properties simultaneously. Along the spectrum, there are methods such as empirical models, semi-analytical models, and subhalo abundance matching techniques (see Ref. [367] for a review).

Ideally, we would like to run a large set of large-volume and high-resolution hydrodynamic simulations for different cosmologies, astrophysics, and DM properties to get accurate theoretical predictions for the abundance and spatial distribution of galaxies/subhalos. However, that is computationally unfeasible at present. For instance, the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS, [171]) and the DaRk mattEr and Astrophysics with Machine learning and Simulations (DREAMS, [172]) projects, which contain thousands of state-of-the-art hydrodynamic simulations that vary cosmology, astrophysics, and DM properties, are two recent attempts to accomplish this but the volume they can sample is relatively small.

On the other hand, HODs are widely used by the community, given that the computational cost of N-body simulations is much smaller than that of their full hydrodynamic

counterparts. In these methods, galaxies are painted on top of the DM halos of N-body simulations, following a set of recipes that aim to capture the main physics describing the halo-galaxy connection. While these methods are designed to match some summary statistics (e.g. galaxy abundance and 2-point correlation function) they may fail for other summary statistics or probes of the entire observed distribution (e.g. field level analyses) as they rely on simplified assumptions about the halo-galaxy connection.

It is therefore desirable to have the best of both approaches: the precision and accuracy of full hydrodynamic simulations at the computational cost of HODs. This is precisely the goal of this chapter: to develop a model that can generate galaxy/subhalo catalogs given the properties of a parent halo and the value of the astrophysical and DM parameters. Our goal is that our model not only accurately matches a given set of summary statistics but instead generates catalogs by sampling the complex and high-dimensional distribution that characterize the properties of subhalos/galaxies such as their positions, velocities, and internal properties (e.g., stellar mass, concentration, etc). We note that recent works have exploited these ideas (e.g., [218], [368]) by either painting galaxy properties on top of subhalos or modeling the displacement vectors between subhalos and galaxies.

In this chapter, we introduce a generative machine learning framework based on normalizing flows [217], [369], [370] and variational diffusion models [233], [235], [371], to model the connection between halos and their subhalo/galaxies at the field level. Distinct from most HODs, our model can populate DM halos with subhalos that are jointly characterized by multiple properties (positions, velocities, and internal properties), taking into account intra-galaxy (i.e., properties of a single galaxy) as well as inter-galaxy (i.e., between galaxies in a catalog) properties.

We train our model using data from the DREAMS project [172] that contains high-resolution zoom-in simulations of Milky Way-like halos. Each of those simulations has a different value of the astrophysical parameters (modeling the strength of supernova and AGN feedback) and has been run with a different warm dark matter (WDM) mass. Once trained, our model can generate a galaxy catalog that represents the subhalos residing in a halo, given some values of the astrophysical parameters and the WDM mass. Our model returns a galaxy catalog with the position, velocity, and internal properties of each subhalo inside the halo. We use multiple summary statistics to quantify the accuracy and precision of the generated galaxy catalogs and compare them with the ones from the simulations. Our model, coined `NeHOD`, exhibits the accuracy of the full hydrodynamic simulations while being orders of magnitude faster. The approach may thus play an important role in studies that require modeling the halo-galaxy/subhalo connection, e.g. galaxy redshift surveys and strong lensing studies.

This chapter is structured as follows. In Section 4.2, we briefly summarize the DREAMS project and the WDM zoom-in suite of simulations used as the training dataset. In Section 4.3, we describe the training dataset in 4.3.1, the machine learning models in 4.3.2, and the training procedure in 4.3.3. In Section 4.4, we show examples of generated subhalo realizations and quantify the fidelity of the trained emulator model. In Section 4.5, we compare `NeHOD` to previous emulator frameworks and discuss its current limitations and future prospects. Finally, we conclude in Section 4.6.

4.2 Simulations

The DREAMS project[1], introduced in [172], is a collection of thousands of state-of-the-art hydrodynamic simulations. These include both cosmological simulations (Uniform Box suites), which provide extensive samples of halos and galaxies, and zoom-in simulations (Milky Way suites) for higher-resolution studies of smaller scales. Similar to the CAMELS project [171], the DREAMS simulations vary both cosmological and astrophysical parameters. On the other hand, DREAMS also vary parameters related to the particle properties of DM for different models (e.g. warm DM, self-interacting DM, etc.). The DREAMS project is the first set of simulations designed to train machine learning algorithms by means of running simulations that sample a large parameter space (cosmology plus astrophysics plus DM properties). Among the simulations currently completed are the WDM Uniform Box and Milky Way suites, with the former explored in [201]. For the full specifications of the simulations, we refer readers to Table 1 of [172].

Here, we briefly describe the WDM Milky Way suite used for the machine learning study in this work. The WDM Milky Way suite contains 1024 simulations of Milky Way-mass halos with varying WDM mass and astrophysical parameters. The simulations are run with the moving-mesh code Arepo [372], which uses the TreePM algorithm to solve the self-gravity magnetohydrodynamic equations [373], [374]. The gravitational softening length is $305\,\mathrm{pc}\,h^{-1}$. The initial conditions are generated by the MUlti-Scale Initial Conditions (MUSIC; [375]) code at $z = 127$ using the second-order Lagrangian perturbation theory (2LPT). Cosmological parameters are fixed at $\Omega_\mathrm{m} = 0.301712$, $\Omega_\Lambda = 0.698288$, $\Omega_\mathrm{b} = 0.046$, $\sigma_8 = 0.839$, $H_0 = 69.1\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ ($h = 0.691$), which is consistent with the Planck 2016 cosmology [376]. The DM and baryon mass resolution are 1.2×10^6 and $1.9 \times 10^5\,\mathrm{M_\odot}h^{-1}$ respectively, which is comparable to the mass resolution of the TNG50 simulation [377], [378], which has a DM and baryon mass resolution of 6.5×10^5 and $1.2 \times 10^6\,\mathrm{M_\odot}h^{-1}$.

The simulations adopt the IllustrisTNG galaxy formation model, as detailed in [40],

[1] https://www.dreams-project.org/

[41]. This model incorporates AGN feedback, self-consistent magnetohydrodynamic, and an updated prescription of stellar formation and evolution, galactic winds, and outflows from the previous Illustris model [379], [380]. Parameters are kept the same as the TNG model, except for two parameters that control supernova (SN) feedback and one that controls AGN feedback. Prescriptions of SN and AGN feedback are currently among the key sources of systematic uncertainty in cosmological simulations. These feedback mechanisms introduce significant challenges in distinguishing the effects of baryonic processes from the inherent properties of DM, e.g. by altering the inner density profiles in dwarf galaxies ([289], [290]; see also [79] for a review). Decoupling baryonic effects from DM properties is one of the primary goals of the DREAMS simulations. For a more detailed discussion of the baryonic prescription in DREAMS, we refer readers to Section 2.1 of [172]. Here, we quickly summarize the three feedback parameters and their relevant physical processes.

The two stellar feedback parameters control the energy and velocity of SN-induced winds, which expel gas from galaxies. In the TNG model, the strength of stellar feedback is quantified using the mass-loading factor:

$$\eta_w \equiv \frac{\dot{M}_w}{\dot{M}_{\rm SFR}} = \frac{2}{v_w^2} e_w (1 - \tau_w), \tag{4.1}$$

where e_w is the specific wind energy in units of $10^{51}\,\mathrm{erg\,M_\odot^{-1}}$, τ_w is the fraction of energy released thermally, and v_w is the wind velocity. The wind velocity depends on the local DM velocity dispersion $\sigma_{\rm DM}$ through the redshift-dependent relation,

$$v_w = \max\left[\kappa_w \sigma_{\rm DM} \left(\frac{H_0}{H(z)}\right)^{1/3}, v_{w,\rm min}\right], \tag{4.2}$$

where κ_w is a dimensionless normalization factor, and $v_{w,\rm min}$ is the minimum wind velocity. In DREAMS, τ_w and $v_{w,\rm min}$ are set to their fiducial values $\tau_w = 0.1$ and $v_{w,\rm min} = 350\,\mathrm{km/s}$, and e_w and κ_w are varied logarithmically over a range of $e_w \in [0.9, 14.4]$ and $\kappa_w \in [3.7, 14.8]$, respectively. The AGN feedback parameter controls the feedback energy released from accretion, which takes the functional form:

$$\Delta\dot{E} = \epsilon_{\rm f,high} \epsilon_r \dot{M}_{\rm BH} c^2, \tag{4.3}$$

where $\epsilon_{\rm f,high}$ is the fraction of energy transferred to nearby gas, ϵ_r is the radiative efficiency, and $\dot{M}_{\rm BH}$ is the black hole accretion rate. $\epsilon_{\rm f,high}$ is varied logarithmically from $[0.025, 0.4]$, and other parameters are kept fixed at their TNG values.

In addition to the three feedback parameters, each simulation in the WDM Milky Way

Table 4.1: Parameter range of the suite of 1024 simulations used in this work.

Parameters	Range	Prior	Fiducial
$M_{\rm WDM}$	$[1.8, 30]$ keV	Inverse Uniform	∞
e_w	$[0.9, 14.4]$	Log Uniform	3.6
κ_w	$[3.7, 14.8]$	Log Uniform	7.4
$\epsilon_{\rm f,high}$	$[0.025, 0.4]$	Log Uniform	0.1

suite also varies over the mass of the WDM particles $M_{\rm WDM}$. The WDM mass is varied inversely from $M_{\rm WDM} \in [1.8, 30]\,{\rm keV}$. These parameters are varied according to a Sobol sequence [381], which ensures a quasi-random, well-distributed sampling across the parameter space that is suitable for machine learning applications. The full prior specification of the WDM mass and feedback parameters, along with their fiducial values, can be found in Table 4.1.

Each simulation zooms in on a high-resolution region that contains a MW-mass galaxy. These galaxies are selected such that they are not in the vicinity of another massive galaxy, i.e. they are not MW-analog. This choice is to reduce the computational cost associated with simulating dense environments. Details on the zoom-in procedure are outlined in Appendix A of [172]. Similarly, as in TNG, the halos and subhalos are identified using the friends-of-friends (FoF) and SUBFIND algorithms [161] with a linking length of $b = 0.2$. We will explore the impact of the choice of the halo finder on the emulated halo and subhalo properties in future work.

4.3 Methodology

In this section, we introduce the `NeHOD` framework. The framework consists of two main components: (1) a conditional normalizing flow for predicting the properties of the Milky Way-mass halo and its central subhalo given the simulation parameters, and (2) a variational diffusion model for predicting the properties of satellite galaxies of the corresponding halo and central subhalo. The framework operates hierarchically, where the diffusion model is conditioned on the output from the normalizing flows (i.e., the halo and central subhalo properties). This hierarchical approach ensures that the dependencies between the properties of a halo, its central subhalo, and its satellites are accurately modeled. A visual representation of the `NeHOD` framework is shown in Figure 4.1.

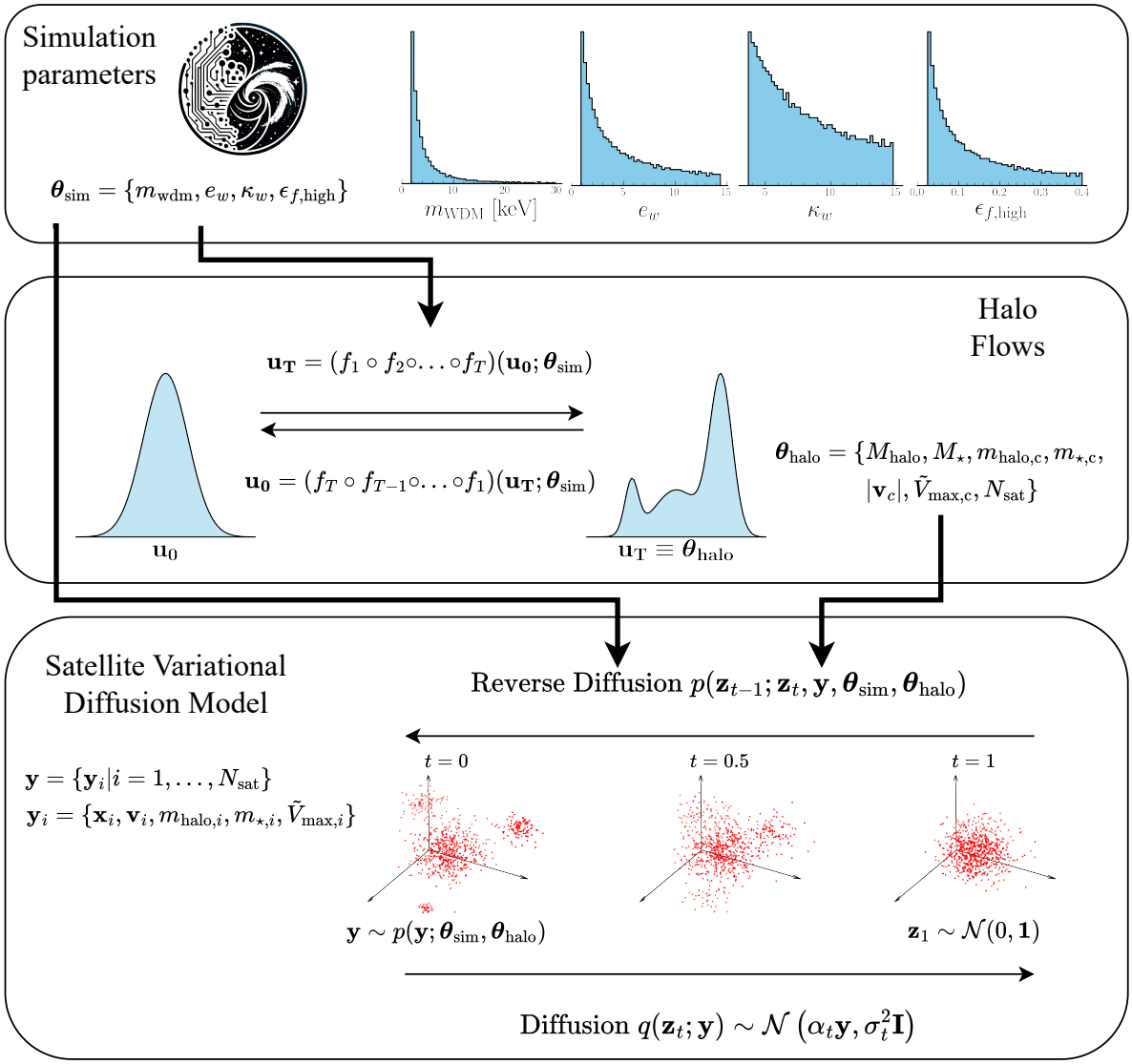

Figure 4.1: Flow chart of the `NeHOD` framework.

In Section 4.3.1, we describe the procedure for creating and preprocessing the dataset as well as the halo properties used as input and output features. We then describe the `NeHOD` architecture and training strategy in Section 4.3.2 and 4.3.3, respectively.

4.3.1 Dataset

For each simulation, we extract the high-resolution halo with approximately MW-mass together with its subhalos from the group catalogs as follows. First, we identify halos with a total mass within $7 \times 10^{11}\,\mathrm{M}_\odot < M_{\rm halo} < 3 \times 10^{12}\,\mathrm{M}_\odot$. Next, we choose the halo with the least contamination from low-resolution DM particles. We require that the total mass of the low-resolution DM particles does not exceed 2% of the total mass within the virial radius in all of our halos. From the 1,024 zoom-in simulations, 1,018 halos satisfy the above criteria.

We now describe the halo and subhalo properties used as input and output features of the `NeHOD` framework. Table 4.2 provides a brief description of the relevant features and the corresponding FoF/Subfind field. As mentioned above, we separate the framework into two components: (1) given the simulation parameters, predicting the properties of the halo and its central subhalo, and (2) given the halo, central subhalo, and simulation parameters, predicting the properties of the satellite subhalos.

Given some feature vector $\boldsymbol{\theta}_{\rm sim}$ (commonly referred to as the "conditioning context") of the simulations, we are interested in emulating the halo and central subhalo. Here, we let the context $\boldsymbol{\theta}_{\rm sim}$ be the WDM mass and the astrophysical parameters in Section 4.2 (i.e., the SN specific wind energy and velocity normalization, AGN parameter),

$$\boldsymbol{\theta}_{\rm sim} = \{M_{\rm WDM}, e_w, \kappa_w, \epsilon_{\rm f,high}\} \in \mathbb{R}^4. \tag{4.4}$$

Note that in practice, we use the inverse of the WDM mass and the logarithm of the feedback parameters because they are sampled uniformly in this space (see Table 4.1). These parameters form a 4-dimensional input vector for our framework. We now describe how our model predicts the properties of the 1) halo, 2) central subhalo, and 3) individual satellite subhalos.

Halo. The conditional normalizing flow takes as input the simulation parameters $\boldsymbol{\theta}_{\rm sim}$ and outputs the properties of the halo and its central subhalo. For the halo properties, we are interested in modeling the halo virial mass $M_{\rm halo}$, total stellar mass $M_\star$, and the number of satellite subhalos $N_{\rm sat}$. The virial mass $M_{\rm halo}$ is estimated as the total mass enclosed within R_{200}, the radius where the average density of the halo is 200 times the critical density of the Universe. We use a reference system where the center of the halo and its peculiar velocity

are both at (0, 0, 0). As mentioned above, we model the central subhalo independently from the satellite subhalos (more details below).

Central subhalo. Our framework models the properties of the central subhalo, $\mathbf{y}_c$, independently of its satellite subhalos. We do this for several reasons. First, as mentioned above, the positions and velocities of the subhalos are in the frame of reference of the halo. Thus, by definition, the position of the central subhalo is always at $\mathbf{x}_c = (0, 0, 0)$. Note that, in general, there may be a small bias between the velocities of the halo and the central subhalo. However, assuming this velocity bias is isotropic[2], we only need to model its magnitude, i.e. $|\mathbf{v}_c|$. Second, the central subhalo dominates the mass within the halo and is significantly more massive than its satellite subhalos, typically by a few orders of magnitude in $M_\odot$. It plays a significant role in the dynamics of satellite subhalos through processes such as tidal stripping, dynamical friction, and mergers, which can alter the mass distribution within the halo and affect the formation and evolution of galaxies within these subhalos [34], [382], [383]. For these reasons, we expect the properties of the central subhalo to vastly differ from those of satellite subhalos. Thus, modeling all subhalos at the same time may be challenging and we decided to model them separately.

In addition to $|\mathbf{v}_c|$, the halo mass $M_{\rm sub,c}$ and the stellar mass $M_{\rm sub,\star,c}$ of subhalos, we are interested in modeling the relation between the total mass and subhalo structural parameters, such as the DM concentration, which has been found to strongly correlate with the halo formation time and environment in many assembly bias studies [e.g., 356]–[359]). Because the DM concentration is not available in SUBFIND, we instead use the quantity $\tilde{V}_{{\rm max},i} \equiv V_{{\rm max},i}/(H_0 r_{{\rm max},i})$, where $V_{{\rm max},i}$ is the maximum circular velocity of the subhalo and $r_{{\rm max},i}$ is the radius at this velocity. This follows from [384], which demonstrates that $\tilde{V}_{\rm max}$ is a good proxy for the DM concentration. Following [385], we can relate $\tilde{V}_{\rm max}$ to the DM concentration c of a Navarro-Frenk-White (NFW, [386]) density profile using the following relations:

$$\delta_v = 2\tilde{V}_{\rm max}^2 = \frac{\delta_c}{7.213}, \text{ where } \delta_c = \frac{200}{3}\frac{c^3}{\ln(1+c) - c/(1+c)}. \tag{4.5}$$

Here, δ_v is the mean overdensity within $r_{\rm max}$, and δ_c is the characteristic NFW overdensity. Note that the relation between $\tilde{V}_{\rm max}$ and c is one-to-one, so both $\tilde{V}_{\rm max}$ and c can be used interchangeably as training features. Here, we choose to use $\tilde{V}_{\rm max}$ instead of c because $V_{\rm max}$ and $r_{\rm max}$ are directly available from the FOF/SUBFIND catalog.

[2]This is a reasonable assumption because the gravitational forces acting upon the central subhalo from various directions tend to average out due to the symmetric distribution of matter within the halo, resulting in an overall isotropic velocity distribution.

To summarize, for the central subhalo, our model predicts these four properties:

$$\mathbf{y}_c = \{M_{\rm sub,c}, M_{\rm sub,\star,c}, |\mathbf{v}_c|, \tilde{V}_{\rm max,c}\} \in \mathbb{R}^4, \tag{4.6}$$

Combining with the halo properties, the output of the conditional normalizing flows can be represented as a vector $\boldsymbol{\theta}_{\rm halo}$:

$$\boldsymbol{\theta}_{\rm halo} = \{M_{\rm halo}, M_\star, M_{\rm sub,c}, M_{\rm sub,\star,c}, |\mathbf{v}_c|, \tilde{V}_{\rm max,c}, N_{\rm sat}\} \in \mathbb{R}^7, \tag{4.7}$$

These features, along with the simulation parameters $\boldsymbol{\theta}_{\rm sim}$, are used as the context of the variational diffusion model for generating the satellite subhalos.

Satellite subhalos. For the features of each satellite subhalo $\mathbf{y}_i$, our model predicts the 3-dimensional positions $\mathbf{x}_i$, velocities $\mathbf{v}_i$, the halo mass $M_{\rm sub,i}$, the stellar mass $M_{\rm sub,\star,i}$, and the concentration proxy $\tilde{V}_{{\rm max},i}$. We only include subhalos with at least 100 DM particles, which corresponds to a minimum mass of $1.2 \times 10^8\,{\rm M}_\odot h^{-1}$. Below this threshold, SUBFIND does not reliably identify subhalos. The features are thus:

$$\mathbf{y}_i = \{\mathbf{x}_i, \mathbf{v}_i, M_{\rm sub,i}, M_{\rm sub,\star,i}, \tilde{V}_{{\rm max},i}\} \in \mathbb{R}^9, \tag{4.8}$$

forming a 9-dimensional output vector for each subhalo. Our framework outputs a vector $\mathbf{y} = \{\mathbf{y}_i | i = 1, ..., N_{\rm sat}\} \in \mathbb{R}^{N_{\rm sat}\times 9}$ containing the properties of $N_{\rm sat}$ subhalos. The number of satellite subhalos $N_{\rm sat}$ is inherently random and varies based on the context $\boldsymbol{\theta}_{\rm sim}$, e.g., a larger value of $M_{\rm WDM}$ will result in fewer subhalos per halo. We will discuss the machine learning architecture suitable for predicting variable numbers of output in Section 4.3.2 below.

Table 4.2: List of parameters and their corresponding FoF/SUBFIND fields used in this work.

Parameters	Description	Relevant FoF/SubFind Field
$M_{\rm WDM}$	Mass of WDM particles	`n/a`
e_w	Specific wind energy of SN wind	`n/a`
κ_w	Dimensionless scaling factor of SN wind velocity	`n/a`
$\epsilon_{\rm f,high}$	Fraction of energy transferred to nearby gas due to AGN accretion	`n/a`
$M_{\rm halo}$	Halo virial mass	`Group_M_Crit200`
$M_\star$	Halo stellar mass	`GroupMassType PartType4`
$\mathbf{x}$	3-dimensional subhalo position vector	`SubhaloPos`
$\mathbf{v}$	3-dimensional subhalo velocity vector	`SubhaloVel`
$M_{\rm sub}$	Satellite subhalo total mass	`SubhaloMass`
$M_{\rm sub,\star}$	Satellite subhalo stellar mass	`SubhaloMassType PartType4`
$V_{\rm max}$	Maximum circular velocity	`SubhaloVmax`
$r_{\rm max}$	Radius at $V(r_{\rm max}) = V_{\rm max}$	`SubhaloVmaxRad`
$\tilde{V}_{\rm max}$	$\tilde{V}_{\rm max} \equiv V_{\rm max}/(H_0 r_{\rm max})$, proxy for DM concentration	`n/a`
$N_{\rm sat}$	Number of satellite subhalos	`n/a`

4.3.2 Machine learning framework

For each set of simulation parameters $\boldsymbol{\theta}_{\rm sim}$, we would like to model and sample from the joint likelihood $p(\mathbf{y}, \boldsymbol{\theta}_{\rm halo} \mid \boldsymbol{\theta}_{\rm sim})$, where $\boldsymbol{\theta}_{\rm halo}$ is the properties of the halo and central subhalo, and $\mathbf{y} = \{\mathbf{y}_i | i = 1, ..., N_{\rm sat}\}$ is a set of interested properties of $N_{\rm sat}$ subhalos. As mentioned previously, since we model $\boldsymbol{\theta}_{\rm halo}$ and $\mathbf{y}$ independently and hierarchically, we can rewrite the final target distribution as:

$$p(\mathbf{y}, \boldsymbol{\theta}_{\rm halo}|\boldsymbol{\theta}_{\rm sim}) = p(\mathbf{y}_1, ..., \mathbf{y}_{N_{\rm sat}}|\boldsymbol{\theta}_{\rm sim}, \boldsymbol{\theta}_{\rm halo}) p(\boldsymbol{\theta}_{\rm halo}|\boldsymbol{\theta}_{\rm sim}). \tag{4.9}$$

This consists of two independent components: a halo likelihood $p(\boldsymbol{\theta}_{\rm halo}|\boldsymbol{\theta}_{\rm sim})$, which will be modeled by the flows, and a satellite likelihood $p(\mathbf{y}_1, ..., \mathbf{y}_{N_{\rm sat}}|\boldsymbol{\theta}_{\rm sim}, \boldsymbol{\theta}_{\rm halo})$, which will be modeled by the variational diffusion model.

Halo likelihood. To model the properties of the halo and central subhalo, we use a normalizing flow [217], [369], [370] that is conditioned on simulation parameters $\boldsymbol{\theta}_{\rm sim}$. Normalizing flows employ non-linear, invertible transformations (with tractable Jacobians and its inverse) from a standard, base Gaussian distribution to a complex target distribution, allowing for efficient density estimation and sampling. They have been employed in various astrophysical applications including generative modeling and simulation-based inference (see [231] for a review). In our framework, our flow model consists of 8 neural spline flows (NSF, [387]) layers, each using a rational quadratic spline transformation with 8 bins (9 knots). To read in the conditioning context, each flow also has a multi-layer perceptron (MLP) with 2 layers of hidden size 16 and `ReLU` activation function. Each MLP layer is followed a batch normalization layer [227] and a dropout layer [226] with a rate of 0.2. We found that due to the limited number of simulations available in our application, batch normalization and dropout are especially important to reduce overfitting in the flows.

Satellite likelihood. To emulate the satellite subhalo population, we adopt an approach similar to [219] and model it as a point cloud, i.e., a set of points in 3-dimensional space with attributes such as velocities, masses, and concentrations. Point clouds eliminate the need for binning and voxelization, which introduces additional hyperparameters that can complicate the training and analysis process. More importantly, this allows us to retain the most information from the data and resolve arbitrary small scales limited only by the spatial resolution of the simulations. Similarly, we also employ a variational diffusion model (VDM, [235], [388], [389]). VDMs are a class of generative models that progressively learn to generate complex data distributions through a process of gradually denoising samples

(referred to as the "reverse diffusion"), starting from a random noise distribution. They offer a few advantages over other generative models. VDMs are more expressive than traditional variational autoencoders, (VAE, [390]), allowing for more flexible network architecture than normalizing flows. Unlike generative adversarial networks,(GANs, [391]), they are generally much easier to train and less prone to experiencing mode collapse [392]. Additionally, they facilitate the tractable evaluation of likelihood, which allows them to be used for inference and anomaly detection.

The VDM consists of a *noise schedule*, $\gamma(t)$, which determines how much Gaussian noise is added to the data $\mathbf{y}$ at each step t of the forward diffusion process, and a *noise prediction model*, $\hat{\boldsymbol{\epsilon}}(\mathbf{z}_t, t)$, which estimates the added noise given the time step t and the noisy data $\mathbf{z}_t$ [233], [235], [371]. In our setup, we use a linear noise schedule with trainable parameters $\eta_{\min}$ and $\eta_{\max}$ that determine the minimum and maximum amount of noise added. For the noise prediction model, we use a Transformer with a fully-connected network for the conditioning embedding layers. Further details on the diffusion process, including the noise schedule, and forward and reverse diffusion process, can be found in Appendix B.2. In the subsequent paragraphs, we discuss the noise prediction model.

The Transformer architecture, first introduced [360], leverages the self-attention mechanism to efficiently process sequential data, such as time series. Self-attention allows transformer-based models to dynamically assess and prioritize the relevance of each part of the input, making them powerful models for many applications ranging from large language modeling [393], [394] to more recent astrophysics applications (see e.g. [219], [395]–[398]). In our framework, we use a transformer with a conditioning-embedding layer. Because each sub-halo population is modeled as a set of points with no inherent ordering, we do not include any position encoding or casual masking layer, making the model permutation-equivariant. During the forward pass, both the input data and the conditioning context are mapped onto a 128-dimensional space using separate fully-connected layers. The outputs of these layers are summed together and subsequently passed through 6 transformer layers, each with 128 hidden channels and 4 attention heads, and a 2-layer MLP with 256 hidden channels and a `GELU` activation function [399]. It is also important to highlight that our neural network architecture does not explicitly incorporate geometrical symmetries, such as rotational and translational invariance. Instead, we apply data augmentation techniques during the training process, as detailed in Section 4.3.3.

Lastly, it is worth noting that graph-based neural networks are also permutation-equivariant and capable of efficiently modeling point cloud cosmological data, as have been shown in recent studies (e.g., [93], [201]–[203], [400]). [219] also demonstrates that using a graph convolutional neural network and k-nearest neighbors graphs can yield comparable results

to transformers. An advantage of using graph neural networks is that the computational cost scales more slowly with the number of point cloud data (i.e. the number of satellite subhalos) increase, as self-attention scales quadratically $\mathcal{O}(N_{\rm sat}^2)$. However, in our application, where the number of points is of order $N_{\rm sat} \sim \mathcal{O}(100)$, the performance difference becomes negligible. In addition, constructing graphs from point clouds introduces additional hyperparameters, e.g., the number of neighbors k. For these reasons, we thus choose a transformer-based architecture for our noise prediction model.

4.3.3 Training and Optimization

The mathematical formulations of normalizing flows and diffusion models are comprehensive and thoroughly documented in the machine learning literature. Thus, we opt to omit their detailed derivations from this section. Instead, we provide summaries of both models, including their optimization objectives, in Appendix B.2 for diffusion models and Appendix B.1 for normalizing flows.

We split the dataset into approximately 800 training and 200 validation simulations. The flows and the VDM are trained independently by performing gradient descent using the AdamW [401], [402] optimizer, with a peak learning rate of 0.0005 and a weight decay coefficient of 0.01. We use a cosine annealing learning rate scheduler [403] with 5000 (100) linear warm-up steps, 10,000 (1000) cosine decay steps, and a batch size $N_{\rm batch}$ of 128 (64) for the VDM (flows). The VDM and flows converge after about 50,000 and 10,000 iterations, which takes approximately 30 minutes in total on an NVIDIA Tesla V100 GPU.

As mentioned in Section 4.3.2, we apply data augmentation to help the model learn the relevant geometrical symmetries, such as translational and rotational invariance. We consider only rotational invariance because the coordinates are always in the frame of the central subhalo by construction. The data augmentation is applied *during training*. In other words, a random rotational matrix is applied to the satellite subhalo coordinates on the fly before being passed through the VDM.

4.4 Results

We present results generating novel satellite populations from our model and perform validation tests by comparing key summary statistics of the populations to those in the DREAMS simulations. To generate a test dataset, we randomly sample the WDM mass and astrophysical parameters from the prior distributions in Table 4.1. We first generate the properties of the halo and central subhalo using the flows, and then use the VDM to generate the subhalo

population given the generated halo and central subhalo. In total, we generate 100,000 halos, which take approximately 40 minutes on a GPU using 1,000 diffusion steps for each sample.

4.4.1 Halo and central subhalo properties

We would like to evaluate how well the conditional flows can recover the properties of Milky Way-mass halos and their central subhalos. We divide the conditioning parameters, i.e. the WDM mass and astrophysical parameters, into bins and calculate the median, 16th, and 84th percentiles of each halo and central subhalo property in each bin. The bins are divided inverse-uniformly for the WDM mass and log-uniformly for the astrophysical parameters (see Table 4.1). Each bin contains about 100 and 10,000 halos for the DREAMS and `NeHOD` samples, respectively.

Figure 4.2 shows variations in each of these properties for both the generated samples and the simulations. The top three rows display the halo properties, including the satellite count, virial mass, and stellar mass. The bottom four rows show the central subhalo properties, including the total mass, stellar mass, velocity, and DM concentration. We compute the concentration using the relations in Equation 4.5. The median, 16th, and 84th percentiles of each bin are shown as the solid and dashed lines respectively. The black lines show the generated samples, whereas the blue lines and shaded regions show the simulations.

We briefly note a few important trends in the simulations. First, decreasing the WDM mass $M_{\rm WDM}$ decreases the satellite count $N_{\rm sat}$. This is expected, as lower $M_{\rm WDM}$ corresponds to a larger free-streaming length, inhibiting the formation of low-mass halos in the early Universe. Other halo and central subhalo properties are relatively invariant to the changes in the WDM mass. However, we note that the halos are selected based on the Milky Way-mass criterion in Section 4.3.1, we do not expect to see much variations in the halo and central subhalo mass. On the other hand, changing either the SN specific wind energy e_w and wind velocity κ_w have pronounced effects on the satellite count, stellar masses, and DM concentrations of central subhalos. The satellite count strongly correlates with the mass-loading factor, which increases with e_w and decreases with κ_w, since higher mass-loading factors lead to more efficient expulsion of gas from subhalos, thereby reducing the number of satellites. Increasing e_w and κ_w reduces the cosmic star formation rate density (SFRD) in the early Universe [40], resulting in lower stellar mass at the present day. The strength of the feedback also leads to decreasing DM concentration in central subhalos due to the expulsion of gas and reduction of potential wells. Similarly to [172], we found that the abundance of satellites appears invariant to changes in the BH feedback factor $\epsilon_{\rm f,high}$, suggesting that BH feedback might not play a significant role in altering the small-scale structure within these

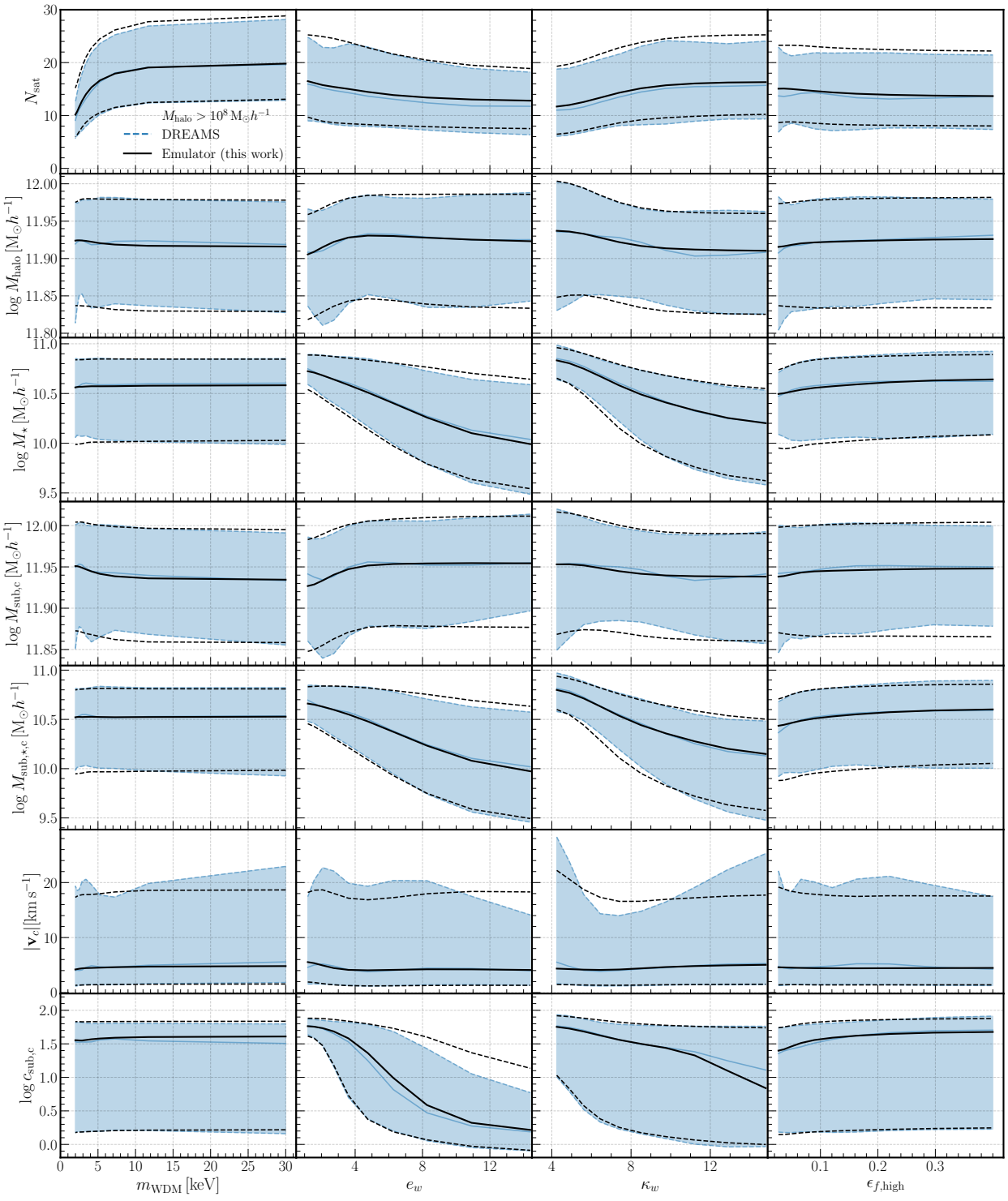

Figure 4.2: The properties of halos and their central subhalo as a function of the WDM mass, SN specific wind energy, SN wind velocity, and AGN parameter (left to right). From top to bottom, the panels show the satellite count, halo virial and stellar mass, central subhalo total mass and stellar mass, central subhalo velocity, and DM concentration. The median (solid lines), 16th, and 84th percentiles (dashed lines) are shown. The black lines denote the samples generated by conditional flows, while the blue lines and shaded regions denote the simulations.

mass ranges in the TNG model.

As evidenced, the halo and central subhalo properties are complex functions of the WDM mass and astrophysical parameters. Therefore, accurately modeling their distributions is crucial for understanding how WDM and baryonic processes influence galaxy formation and evolution. Moreover, in our framework, these parameters are fed directly into the VDM model as conditioning parameters, making the overall performance heavily reliant on the accurate representation of these input parameters (as we will discuss in Section 4.4.2). We now compare the generated samples and the DREAMS samples to evaluate the performance of the conditional flows.

In general, the median halo and central subhalo properties predicted by the flows align well with the simulations. There are minor discrepancies in the median DM concentration of the central subhalos $c_{\rm sub,c}$, though they are small compared to the intrinsic spread of the distribution. We also note that the distribution of $c_{\rm sub,c}$ is bimodal, making it more challenging to capture accurately. Though the origin of this bi-modality is unknown, we find that the peaks are strongly correlated with the stellar mass and the SN feedback parameters e_w and κ_w, suggesting that SN feedback can disrupt DM potential wells in distinct ways. This origin will be the subject of future exploration. We present additional results and analysis in Appendix B.3.1. Compared to the median, the 16th and 84th percentiles are noticeably noisier, and the agreement between the flow samples and the simulations is less robust. In the case of the satellite count $N_{\rm sat}$, the flows slightly overestimate the 84th percentile. The velocity distribution $|\mathbf{v}_c|$ has a long tail at the high end, which is also not perfectly captured by the flows. Similarly to the median, the DM concentration also shows less agreement compared to other properties. Despite these issues, the overall agreement between the generated and simulated properties remains quite good, demonstrating the effectiveness of the conditional flows in modeling halos and subhalos. We expect that with more simulated data, the accuracy of the distributions will further improve.

We note that [172] also presented a machine learning-based emulator for the WDM satellite count (Section 3). Given the four simulation parameters (i.e., the WDM mass and astrophysical parameters), [172] assumes a Gaussian distribution of satellite count and utilizes a fully-connected neural network to predict the mean and standard deviation. Similarly to the conditional flows, this approach also allows both efficient sampling and likelihood evaluation and achieves somewhat similar results to the flows. However, in our framework, we opt to use the flow as it requires fewer assumptions on the truth distributions and thus can model more complex distributions. As seen in Figure 4.2, some properties display non-Gaussian features, such as the long tails in the distributions of the halo and central subhalo stellar mass, central subhalo velocity, and DM concentration. This is especially true when

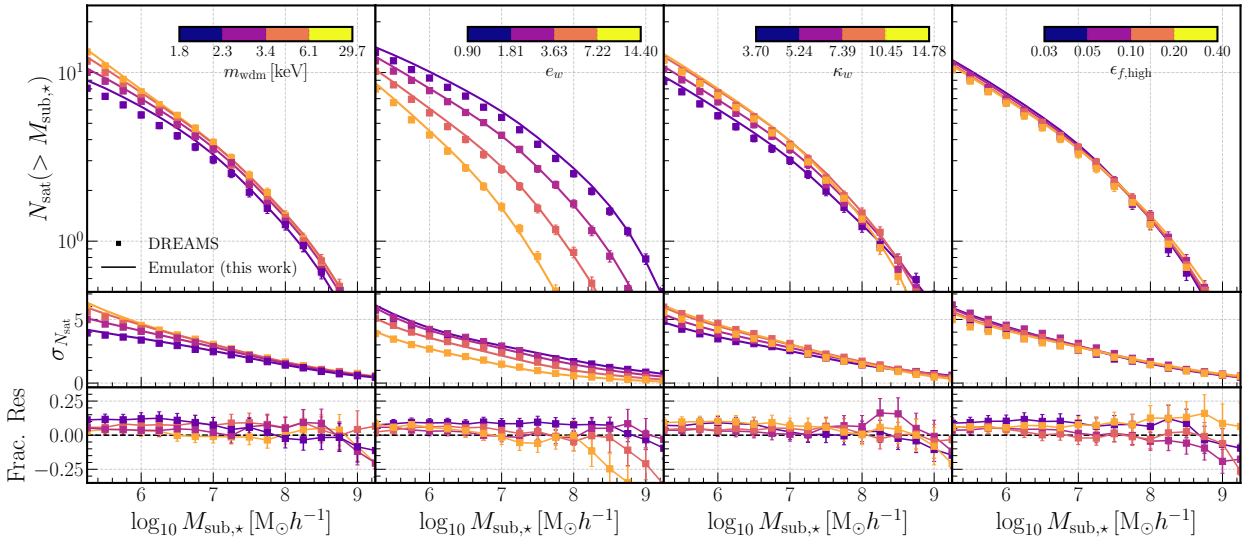

Figure 4.3: *Top*: The satellite stellar mass functions as emulated by `NeHOD` and from the DREAMS simulations. The columns show the variations of the mass functions over the WDM mass and astrophysical parameters. In each column, the corresponding parameter is divided into four bins with approximately the same number of halos, which are approximately 25,000 and 250 for the emulated and simulated halos, respectively. The color denotes the average mass functions for the emulated (solid lines) and the simulated (squares) satellite populations in the corresponding bin. The error bars denote the error in estimating the average mass functions of the simulations, due to the limited number of samples. *Middle*: The standard deviation of the satellite stellar mass functions in. *Bottom*: The residuals of the median stellar mass functions between `NeHOD` and DREAMS.

modeling multi-dimensional distribution, where correlations between halo properties can be non-Gaussian and complex. For example, we expect the total mass and stellar mass of the halos and central subhalos to be strongly correlated though this is not shown directly in Figure 4.2. We further explore these correlations in Appendix B.3.1.

4.4.2 Satellite properties

We now evaluate the properties of the satellite populations generated by the VDM, by comparing key summary statistics. We divide each parameter into four bins such that each bin contains approximately the same number of simulations. Because of the prior distributions in the simulation parameters (as shown in Table 4.1), the bins are distributed inverse-uniformly in WDM mass and log-uniformly for the astrophysical parameters. Each bin contains approximately 25,000 and 250 emulated and simulated satellite populations, respectively. For each bin, we calculate and compare the halo and stellar mass functions (Section 4.4.2), stellar-to-halo mass relations (Section 4.4.2), concentration-mass relations

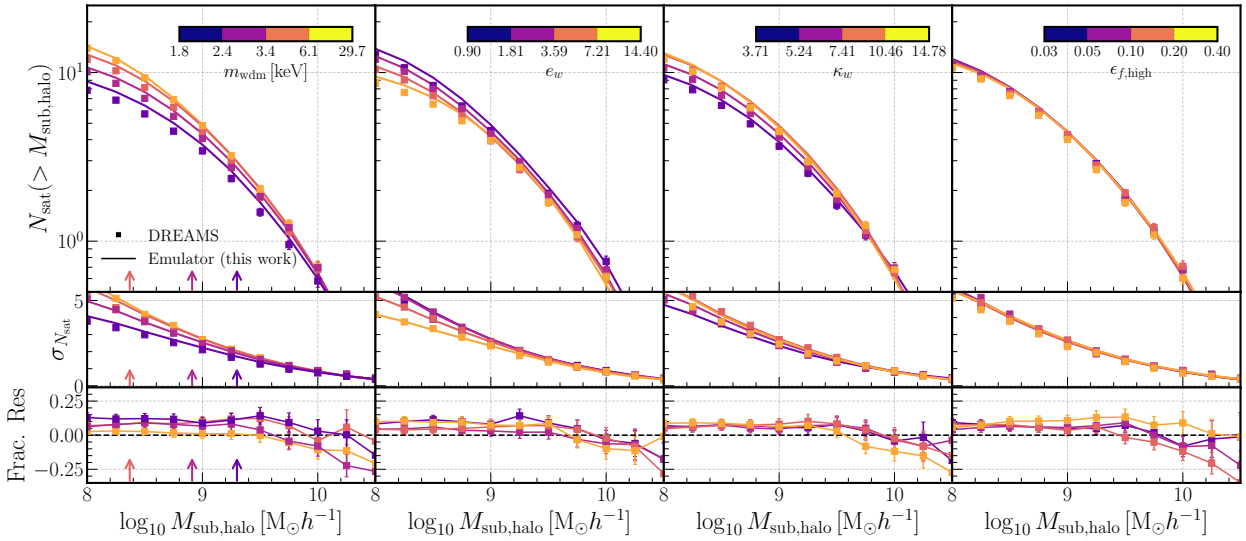

Figure 4.4: The satellite halo mass functions as emulated by `NeHOD` and from the DREAMS simulations. Panels are similar to Figure 4.3. The vertical arrows in the first column (WDM mass) show the half-mode mass $M_{\rm hm}$.

(Section 4.4.2), and position and velocity clustering (Section 4.4.2).

Mass functions

The mass functions are among the key observables that have been used to place constraints on the WDM mass [e.g., 404], [405]. As also discussed in Section 4.4.1 with the abundance of satellites, lower WDM masses lead to fewer low-mass subhalos in the mass functions due to free-streaming effects. The satellite stellar and halo mass functions are thus sensitive to changes in the WDM mass, making them effective tools for constraining DM properties on these scales. However, baryonic feedback can also disrupt the formation of low-mass satellites and leads to similar effects in the mass functions [e.g., 121], [406]. It is thus important to differentiate the effects of DM properties from those of baryonic feedback when interpreting the mass functions.

We show the satellite stellar and halo mass functions respectively in Figures 4.3 and 4.4. From left to right, the columns show variations over the WDM mass $M_{\rm WDM}$, SN specific wind energy e_w, SN wind velocity normalization κ_w, and BH feedback factor $\epsilon_{\rm f,high}$. The top, middle, and bottom panels show the average, standard deviation, and fractional residuals between the generated samples and those from DREAMS, respectively, with each color representing a different parameter bin. The fractional residuals are defined as the differences between the average mass functions divided by the DREAMS mass functions. The error bars denote the standard *errors* of the simulations due to the limited number of samples. In

the WDM panels of Figures 4.4, we also include the WDM half-mode mass $M_{\rm hm}$ as vertical arrows. The half-mode mass $M_{\rm hm}$ characterizes the length scale at which the WDM power spectrum drops by one-half relative to the CDM power spectrum, using the fitting formula from [47], [63]. For our choice of cosmological parameters ($\Omega_{\rm m} = 0.301712$, $h = 0.691$), $M_{\rm hm}$ follows the scaling relation:

$$M_{\rm hm} = 3.64 \times 10^8 \left(\frac{M_{\rm WDM}}{3.0\,\mathrm{keV}} \right)^{-3.33} \mathrm{M}_\odot h^{-1}. \tag{4.10}$$

We expect properties such as the WDM mass functions to deviate from CDM below approximately $M_{\rm hm}$.

The average trends of the mass functions provided by the DREAMS simulations align well with theoretical predictions. In the first column of both Figures 4.3 and 4.4, we observe that decreasing the WDM mass leads to a generally smaller number of low-mass subhalos and shallower mass functions due to the free-streaming effect. On the other hand, changing the SN specific wind energy e_w and SN wind velocity normalization κ_w does not significantly affect the shape of the halo mass functions. The offsets between the bins are mainly due to differences in the number of subhalos, as shown in Figure 4.2. Increasing either the SN specific wind energy or wind velocity normalization results in satellites with lower stellar mass, though the impact of each parameter varies. For example, the effect of SN wind velocity normalization is more subtle and only becomes significant in more massive satellites, potentially due to the wind velocity cap, as discussed in Section 4.4.1. Similar to the halo properties, we find that the BH feedback factor $\epsilon_{\rm f,high}$ does not significantly affect the mass functions.

These differences in parameter trends further highlight the complexity of baryonic feedback processes in shaping the satellite populations. Accurately modeling these processes is thus crucial for addressing parameter degeneracy and interpreting the mass functions. Overall, the VDM can accurately reproduce these trends in both the average and standard deviations of mass functions. However, there are slight offsets at the 10-15% level between the generated mass functions and the simulations in both the stellar and halo mass functions. This discrepancy potentially arises because the conditional flows overestimate the number of satellite subhalos, as illustrated in Figure 4.2. To address this, we investigate the normalized stellar and halo mass functions and find that the normalized mass functions generally provide a better match. We present these results and additional analysis in Appendix B.3.2. Additionally, in both the stellar and halo mass functions, we see that the VDM tends to under-predict the number of massive satellite subhalos, starting at about $M_{\rm sub,\star} \gtrsim 10^{8.5}\,\mathrm{M}_\odot h^{-1}$ and $M_{\rm sub,halo} \gtrsim 10^{9.5}\,\mathrm{M}_\odot h^{-1}$. These differences are generally within

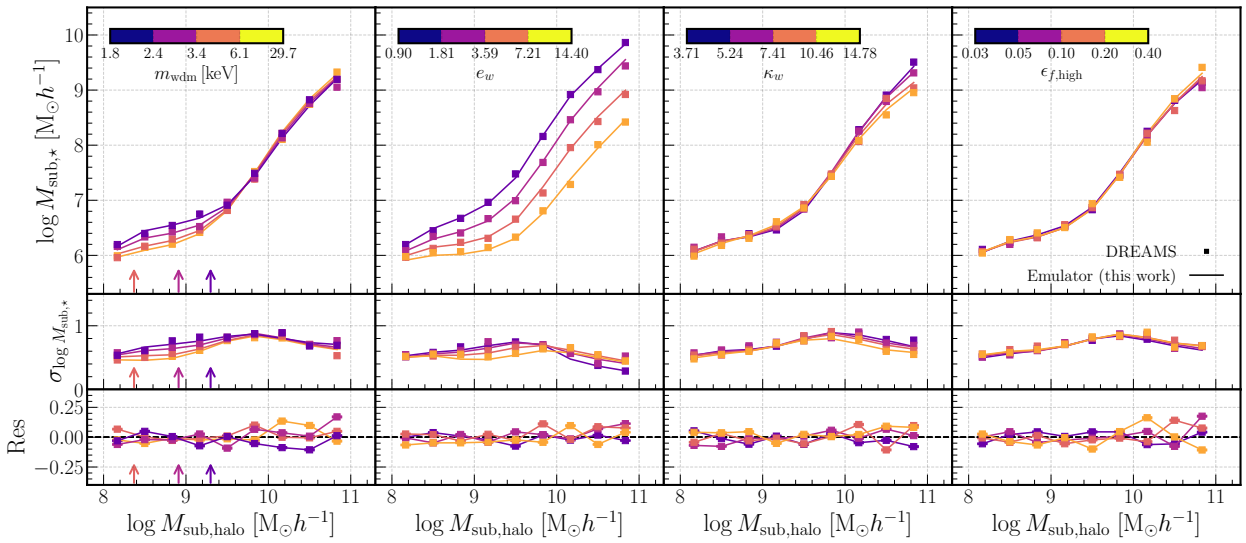

Figure 4.5: The stellar-to-halo mass relations (SHMR) of satellite halos generated by `NeHOD` and extracted from the DREAMS simulations. From the top to bottom panels, we show the averages, standard deviations, and residuals between the `NeHOD` samples and simulations of the SHMR. Each column shows the variations of the SHMR to the WDM mass and astrophysical parameters, with each color denoting a different bin of the corresponding parameter. The error bars denote the errors in estimating the average SHMR of the simulations, due to the limited number of samples. The vertical arrows in the first column (WDM mass) show the half-mode mass $M_{\rm hm}$.

1–2 error bars of the simulations, which also increase significantly towards the high mass end. This is to be expected, as the number of subhalos satisfying $M_{\rm sub,\star} \gtrsim 10^{8.5}\,\mathrm{M}_\odot h^{-1}$ and $M_{\rm sub} \gtrsim 10^{9.8}\mathrm{M}_\odot h^{-1}$ constitute only about 6% and 14% of the subhalos in the simulations, respectively. This rarity also creates challenges for the VDM in learning the distribution of their properties accurately due to the limited representation of these high-mass subhalos in the training data.

Stellar-to-halo mass relation

In addition to the mass functions, we would like to test how well the VDM can capture relationships between the satellite halo and stellar mass. We thus compute the stellar-to-halo mass relation (SHMR), which is a key summary statistic used in many galaxy formation studies (e.g., [271], [407], [408], etc.). The SHMR provides insights into the efficiency of star formation and the impact of feedback processes and thus is sensitive to baryonic prescriptions of the simulations. Accurate modeling of the SHMR is important for understanding galaxy evolution and the role of different baryonic processes in shaping satellite populations.

In Figure 4.5, we compare the SHMRs of the `NeHOD` and DREAMS samples for different

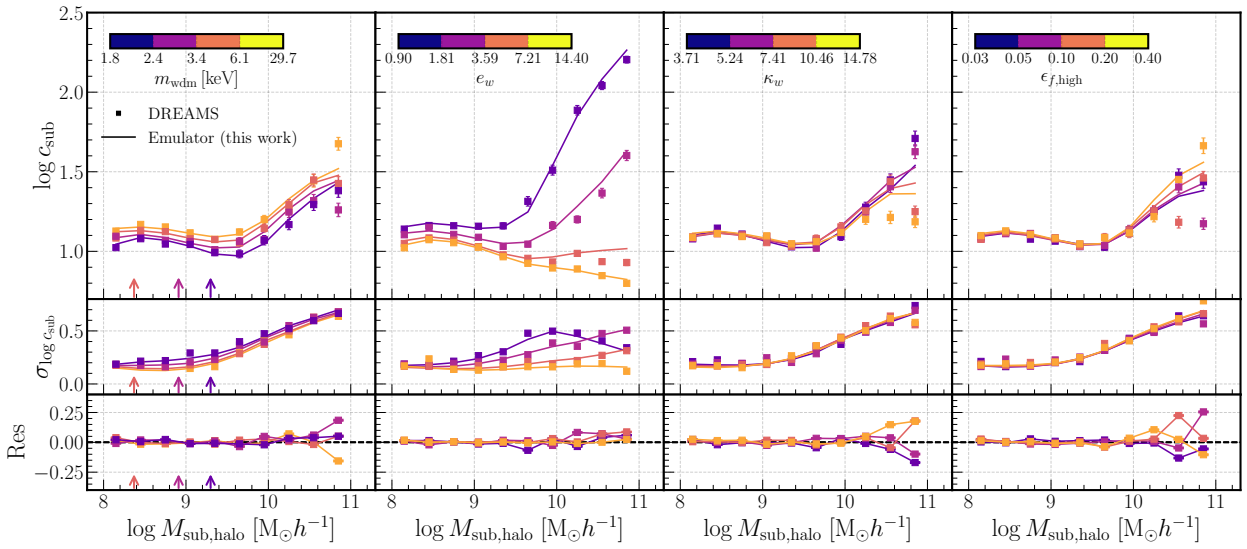

Figure 4.6: The concentration-mass relations of satellite halos generated by `NeHOD` and extracted from the DREAMS simulations. Panels are similar to Figure 4.5.

parameter bins. Similarly to Figures 4.3 and 4.4, we show the average (top), standard deviation (middle), and fractional residuals (bottom) of the SHMR. As expected, the SHMR is sensitive to changes in the SN specific wind energy e_w and wind velocity normalization κ_w, as shown in the second and third columns of Figure 4.5. The latter affects only high-mass subhalos, possibly because of the cut-off in minimum wind velocity in Equation 4.2 and as also discussed previously. These trends in the SHMR also help explain the shape of the stellar mass functions in Figure 4.3. Again, we do not see significant impacts of the BH feedback factor $\epsilon_{\rm f,high}$ at these mass scales. Though somewhat interestingly, for low-mass subhalos ($M_{\rm sub,halo} \lesssim 10^9\,{\rm M}_\odot h^{-1}$), increasing the WDM mass $M_{\rm WDM}$ somewhat increases the subhalo stellar mass. Similar effects of the WDM mass on the baryonic contents of galaxies have also been observed in previous studies (e.g., [407], [409]). A potential explanation could be that WDM delays star formation, resulting in a more extended period of gas availability for star formation. However, investigating these trends further is beyond the scope of this chapter and is left for future work.

Compared to the mass functions, we see a much better agreement between the average SHMRs of the `NeHOD` and DREAMS samples. Deviations from the simulations are typically at about 1-2%, as observed from the bottom panels of Figure 4.5. Likewise, the standard deviations of the SHMR are well-recovered by the VDM. Additionally, we do not observe any significant bias in the SHMR, including in high-mass subhalos. This indicates that the VDM can learn the SHMR well even given the limited representations of these high-mass subhalos, contrary to the mass functions in Figures 4.3 and 4.4. We offer a few potential explanations

for this discrepancy. First, to accurately capture the mass functions, the VDM must learn the correlations between the mass of all satellite subhalos in a given halo. If the number of satellite subhalos is high, this can be difficult to learn. In contrast, the SHMR only requires learning correlations between properties of a single subhalo, particularly the total and stellar mass, which are easier to capture. Additionally, the strong correlations between the stellar and halo mass can facilitate learning and result in a better fit with fewer subhalos. Lastly, as we sample the halo properties using the conditional flows in Section 4.4.1, this can create a mismatch between the halo populations of the `NeHOD` and DREAMS samples if the flows do not capture the halo population perfectly. In the case of the number of subhalos, we see that this can create a slight offset between the average mass functions of the `NeHOD` samples and the simulations. The SHMR might be more invariant to the halo properties such as the number of subhalos, leading to better agreement in the fit.

Concentration-mass relation

We now test how well the VDM can recover the DM concentrations of satellite subhalos. The DM concentration describes the internal structure of a halo and is known to correlate strongly with both the formation time and the environment of the halo [356]–[359]. A key relation in galaxy formation studies is the concentration-mass relation, which links the DM concentration to the halo mass.

Using Equation 4.5, we convert $\tilde{V}_{\rm max}$ to DM concentration $c_{\rm sub}$ and calculate the concentration-mass relations for `NeHOD` samples and the DREAMS simulations. Figure 4.6 presents the concentration-mass relations for the same WDM and astrophysical parameters bins. Similarly, as with the SHMR, we show the average, standard deviation, and fractional residual between the generated samples and simulations. We note that as in the case with the central subhalo, the distribution of $c_{\rm sub}$ is bimodal and strongly correlates with the subhalo stellar mass and SN feedback parameters (i.e., the specific wind energy e_w and velocity κ_w). However, the bi-modality is much less prominent for the satellite subhalos, possibly due to lower stellar mass, making average and standard deviation acceptable summaries for the concentration-mass relations. We present additional results on the DM concentration in Appendix B.3.3.

We highlight several important trends in the concentration-mass relations. Satellite subhalos in simulations with lower WDM mass tend to be slightly less concentrated, as expected since lower WDM mass leads to a suppression of small-scale structure formation and results in less dense subhalos. For low-mass subhalos ($M_{\rm sub} \lesssim 10^9\,{\rm M}_\odot h^{-1}$), the concentration remains relatively constant regardless of variations in WDM mass or feedback parameters. However, for high-mass subhalos, the concentration exhibits significant variability. This is

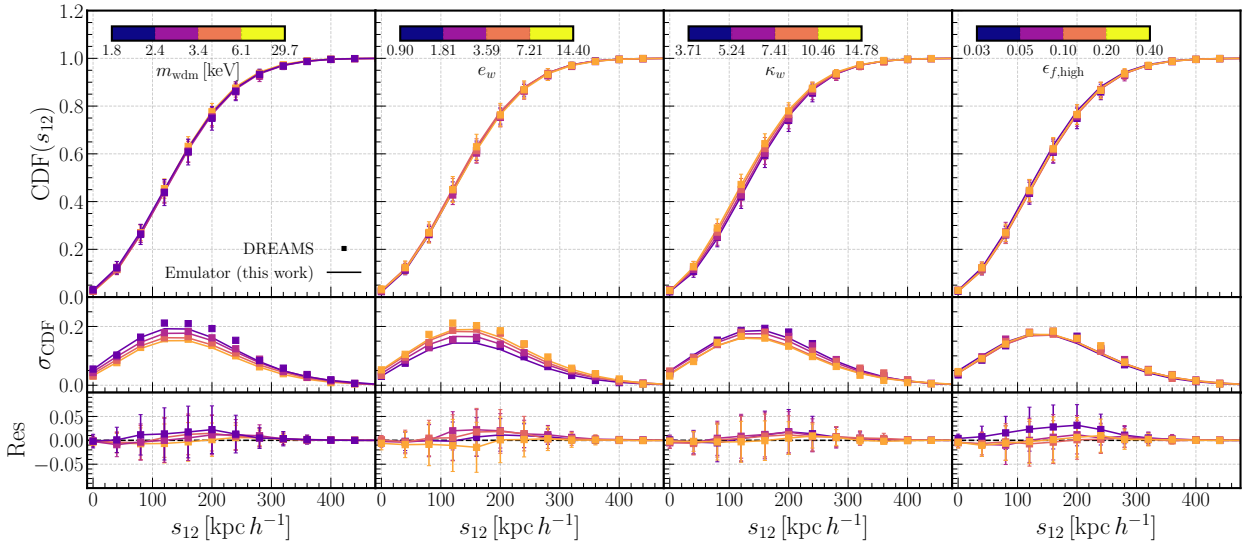

Figure 4.7: The CDFs of the pairwise distances. From top to bottom, the rows show the average CDF, its standard deviation, and the residual between the NeHOD samples and the simulations. Each column shows the variations of the CDF to the WDM mass and astrophysical parameters. Each color shows a different parameter bin. The error bars denote the errors in estimating the CDF of the simulations, due to the limited number of samples.

especially true in the case of the SN specific wind energy e_w, where we see the concentrations of the highest-mass bins ($M_{\rm sub} \approx 10^{11}\,{\rm M}_\odot h^{-1}$) can differ by $2-3$ dex between high and low values of e_w. We also see smaller but notable effects from the SN wind velocity normalization κ_w and the BH feedback factor $\epsilon_{\rm f,high}$. However, the simulation data becomes increasingly noisy at higher masses, as we marginalize over e_w, where the DM concentration varies over a wide range, making it difficult to draw conclusive statements.

In Figure 4.6, the averages and the standard deviations of the concentration-mass relation predicted by NeHOD generally agree with the simulations. We do not observe any significant bias in the overall trends, though there are some notable deviations in the average relations at higher masses in the κ_w and $\epsilon_{\rm f,high}$ panels. As mentioned previously, the concentrations of these high-mass subhalos can vary significantly, resulting in noisier simulation data that can complicate the learning process and make it challenging to discern any clear trend. Despite this, the VDM predicts slight variations in the DM concentrations with κ_w and $\epsilon_{\rm f,high}$.

Position and velocity clustering

As discussed in Section 4.3, the NeHOD framework is unique such that the point-cloud approach allows for more natural modeling of the subhalos phase-space. More importantly, this eliminates reliance on binning or voxelization and can resolve arbitrary small scales,

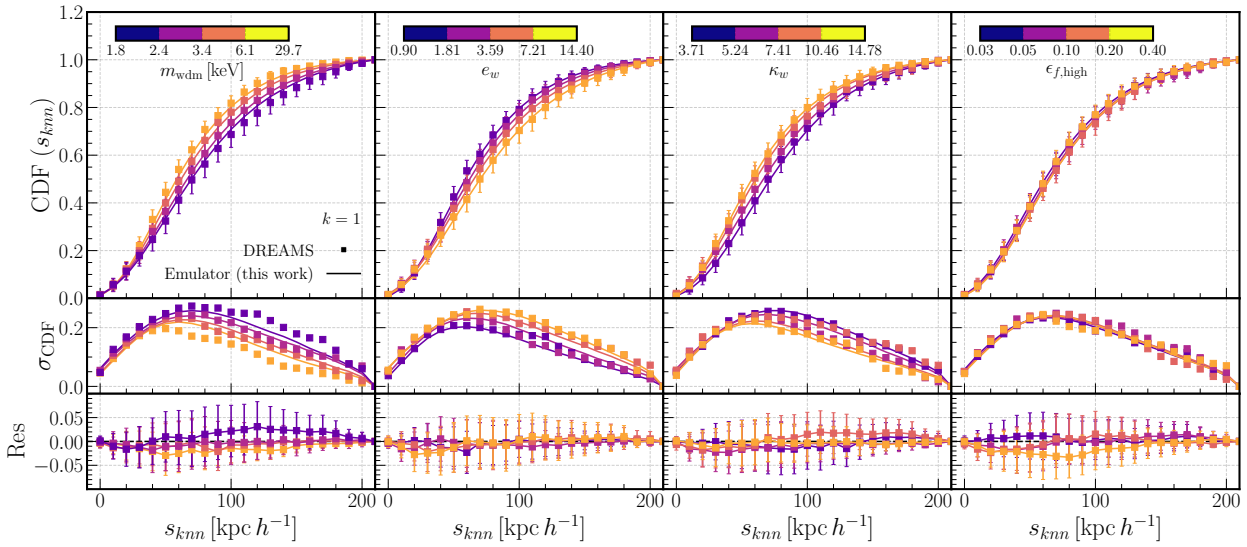

Figure 4.8: The CDFs of the k-nearest neighbor distance for $k = 1$. Panels are the same as Figure 4.7.

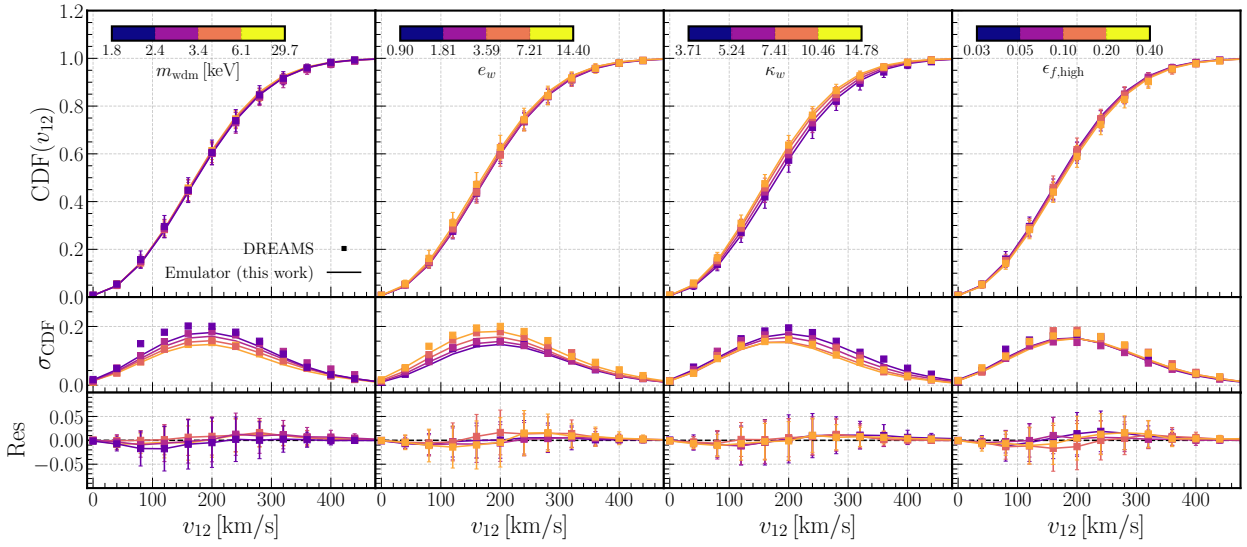

Figure 4.9: The CDFs of the pairwise velocities. Panels are the same with Figure 4.7.

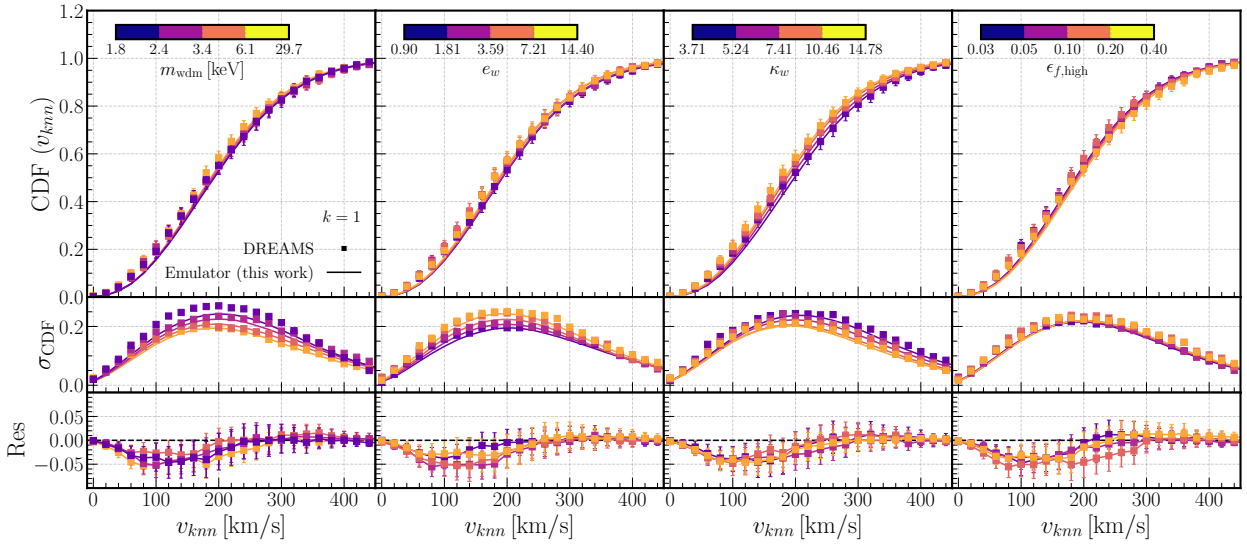

Figure 4.10: The CDFs of the k-nearest neighbor velocities for $k = 1$. Panels are the same with Figure 4.7.

down to the spatial resolution of the simulations. In this section, we examine the clustering of satellites in this space, which is important for applications such as strong lensing and cosmological inference using large-scale structures. We thus compute the pairwise distances in position and velocity space. In addition, we calculate k-nearest neighbor statistics for each satellite, such as the mean distance and velocity for the k nearest satellites.

We first examine the position clustering. Figures 4.7 and 4.8 show the cumulative distribution functions (CDFs) of the pairwise distances and the k-nearest neighbor distance, respectively. From the top to bottom panels, we display the averages, standard deviations, and residuals between the generated samples and the simulations for the CDFs. Figure 4.8 show the k-nearest neighbor distance for $k = 1$, though we observe similar performance for $k = 2$ and $k = 4$ (which we show in Appendix B.3.4).

As shown in Figure 4.7, variations in the WDM mass and astrophysical parameters do not influence the average pairwise distances CDF, though there are notable effects on the spread. This is likely because the pairwise distance distribution is dominated by large-scale structure, which is less sensitive to the details of small-scale physics governed by the WDM mass and astrophysical feedback processes. On the other hand, there are clear variations in both the average and spread of the CDF in Figure 4.8, which is more sensitive to clustering at small scales. As expected, increasing the WDM mass $M_{\rm WDM}$ would result in a steeper CDF (more clustered). Additionally increasing the specific wind energy κ_w and decreasing the wind velocity normalization e_w both result in a steep distribution, as they both increase the SN mass-loading factor and result in a more efficient feedback effect.

In general, we observe that the VDM accurately predicts the above trends in both the pairwise distances and the k-nearest neighbor distance. The average and standard deviation of the CDFs are recovered, though the averages often agree slightly better. This indicates that the VDM can effectively learn the position-space distribution of galaxies. This is significant as it demonstrates the potential of `NeHOD` to capture complex spatial relationships in cosmological data, enabling more precise predictions of galaxy clustering and improving our understanding of the underlying physical processes driving structure formation in the universe.

We now examine the velocity clustering. Figures 4.9 and 4.10 show the averages, standard deviations, and residuals CDFs of the pairwise velocities and k-nearest neighbor velocities. As evidenced, the WDM mass and astrophysical parameters have more subtle impacts on the CDFs of the pairwise velocities and k-nearest neighbor velocities, mainly affecting the spread of the CDFs.

In the case of the pairwise velocities, the VDM can predict both the average CDFs and the standard deviations well for each parameter bin. The VDM captures subtle variations in the CDFs due to changes in the SN wind velocity normalization κ_w, indicating its sensitivity to this parameter. However, for the k-nearest neighbor velocities, while the predicted trends generally align with the simulations, the VDM underestimates the average CDFs at low velocities by about 5%. This trend is observed across all parameter bins, suggesting a systematic shift in the k-nearest neighbor velocities towards high velocities. Though small, this discrepancy suggests that while the VDM effectively captures the general trends in velocity clustering, it may have limitations in accurately modeling the detailed distribution. These findings highlight the challenges in modeling velocity clustering, especially at small scales where various processes and interactions play significant roles. Further refinement of the VDM, possibly incorporating more detailed physical processes or higher resolution data, could improve its predictive accuracy. Understanding these velocity distributions is crucial, as they provide insights into the dynamic state of galaxies and influence various processes such as accretion and tidal interactions.

4.5 Discussion

4.5.1 Comparison with previous works

Emulators such as `NeHOD` can considered as non-parametric techniques for modeling the galaxy-connection simulations. Similar non-parametric, machine learning-based techniques have also been employed in various aspects of galaxy formation and cosmology [e.g., 250],

[406], [410], [411]. The NeHOD framework carries a few notable advantages compared to traditional parametric modeling:

- NeHOD requires fewer assumptions on the distributions of halo and galaxy properties. This is particularly important when such distribution displays non-Gaussian features that are difficult to model. For example, the DM concentration of subhalos in our simulations is bimodal, with this bi-modality correlating non-trivially with the subhalo stellar mass and the stellar feedback parameters (i.e., e_w and κ_w). NeHOD can capture these features, as shown in Appendix B.3.1 and B.3.3.

- Both the VDM and the normalizing flows employed in NeHOD can effectively model high-dimensional distributions. This allows NeHOD to capture the complex relationships between halo and galaxy properties with few assumptions. For example, traditional HOD techniques must assume a functional form for the average SHMR and its scatter to model the stellar and halo mass of subhalos. These assumptions can be further complicated by dependencies on the WDM mass and astrophysical parameters, as shown in Section 4.4.2. Moreover, traditional methods do not scale well with increasing parameter dimensions, as each additional property requires new assumptions. In contrast, NeHOD learns these complex relationships directly from the data, allowing it to handle a larger parameter space more efficiently and accurately.

We also highlight the key advantages of NeHOD over past machine learning-based emulators for modeling galaxy-halo connections.

- Existing emulators have employed techniques such as Gaussian Process for emulating summary statistics like satellite counts and mass functions. In contrast, NeHOD uses a transformed-based VDM that can model individual satellites, which we then extract summary statistics as shown in Section 4.4. This capability provides access to lower-level information that can be leveraged for various applications, such as generating mock subhalo catalogs.

- As discussed in Section 4.3, we represent satellite subhalos as point clouds. In contrast to binning and voxelization, point clouds carry more information about the data and can resolve smaller scales, down to the spatial resolution of the simulations. We see in Section 4.4.2 that NeHOD can successfully capture the general trends in the position and velocity clustering, in particular the CDFs of the pairwise separation and the k-nearest neighbor statistics. Additionally, though not explored in this work, point clouds facilitate the incorporation of symmetry into the model, which can further enhance its accuracy and performance in capturing the underlying physical processes.

- Using VDMs for learning point-cloud data has notable advantages compared to other generative machine learning models. Compared to GANs, VDMs are more stable and less prone to mode collapse. They also provide a tractable way to evaluate the likelihood, making them suitable for inference and anomaly detection tasks. Additionally, VDM is more expressive than VAE and normalizing flows. Compared to normalizing flows, VDMs also scale better with high-dimensional data.

4.5.2 Current limitations and future outlook

We now discuss the current limitations and future prospects of the `NeHOD` framework.

- Although our framework can recover the general trends in the simulations, it tends to struggle at the extreme values. In particular, as shown in Appendix B.3.2 and Figures 4.3 and 4.4, the model tends to under-predict the number of massive subhalos in the stellar and halo subhalo mass functions. This issue likely arises from the rarity of these subhalos and the limited number of training simulations available. We anticipate that this limitation will be mitigated as we incorporate more simulations into our training dataset.

- In Section 4.4.2, we compare the clustering in position and velocity space between the `NeHOD`-generated samples and the simulations by computing the pairwise separations and k-nearest neighbor statistics. The position clustering is well-recovered; however, we observe that the distributions of the k-nearest neighbor velocities predicted by `NeHOD` are slightly shifted towards the high velocities. Understanding the local velocity distributions is important to capture physical processes such as accretion and tidal interactions. In future work, we will focus on refining the model to correct this velocity shift by incorporating additional velocity constraints and enhancing the training data with more diverse simulation scenarios.

- In this work, our model is conditioned on the properties of the central subhalo of the considered halo from the hydrodynamic simulation. Ideally, we would like to condition our model on the properties of the central subhalo of the corresponding N-body simulation. We note that in this case, the positions of the halos in the N-body and hydrodynamic simulations may be different, so our model should also predict the displacement vector. We leave this for future work.

- In this work, we have trained our model with halos whose mass lies in a relatively small range. It would be good to train the models on halos with much wider mass variations to fully exploit our method's capabilities.

- Although we apply data augmentation techniques to incorporate SE(3) symmetry, such as rotational equivariance, during training, our model does not explicitly impose these symmetries. It is expected that building such symmetries into our model may improve its performance.

- In this work, the flows model the *joint distribution* of the halo and central subhalo properties. Another interesting avenue for exploration is using a hierarchical normalizing flows approach, similar to what is described in [218]. Specifically, we can employ one flow to model the conditional distribution of halo properties given the simulation parameters, and another to model the conditional distribution of central subhalo properties given the halo properties and simulation parameters. This can be helpful for applications in which we already have access to the halo properties but not the subhalos, e.g., "painting" subhalos onto existing halos in N-body simulations.

4.6 Conclusion

In this work, we develop an emulator for Milky Way-mass halos and their subhalos using normalizing flows and variational diffusion models. Our framework, `NeHOD`, is hierarchical: the normalizing flows capture the conditional distribution of halo and central subhalo properties given the simulation parameters. The diffusion models then learn the properties of satellite subhalos given the halo properties and simulation parameters. This approach allows us to capture the complex dependencies between different levels of structure in the simulations. We represent satellite subhalos of each simulation as a point cloud, which eliminates the need for binning and voxelization and can capture simulation details at extremely fine spatial scales. To model this point cloud, we employ a transformer-based architecture in the diffusion models, which can effectively analyze the intricate spatial relationships and high-dimensional characteristics of the data. This architecture leverages self-attention mechanisms to dynamically focus on different parts of the point cloud, improving the model's ability to learn and represent complex structures.

We train the `NeHOD` framework on the Milky Way WDM simulation suite of the DREAMS project, which includes 1024 simulations with varying WDM mass and astrophysical parameters, and summarize our main results below:

- Using the conditional flows, we sample the halo and central subhalo properties given the WDM mass and astrophysical parameters, and compare the generated samples with the simulations in Section 4.4.1. We show that the flows can capture the median, 16th, and 84th percentiles of the distributions of these properties and their complex

dependencies on the simulation parameters. There are some slight discrepancies in the 84th percentiles of the satellite count and the concentration, though we expect this to improve with more training simulations.

- For each halo and central subhalo generated by the flows, we use the VDM to generate its satellite populations, including the positions, velocities, halo and stellar mass, and DM concentration of each satellite. In Section 4.4.2, we extract key summary statistics, including the stellar and halo mass functions, stellar-to-halo mass relations, concentration-mass relations, and spatial and velocity clustering. We demonstrate that the VDM can generally capture these summary statistics. We observe small disagreements in the outer bins of the mass functions, which could be attributed to the limited number of high-mass satellites in the training data. Additionally, there is a slight shift towards higher velocities in the k-nearest neighbor velocity distributions, although the overall trends remain consistent. In future work, we will focus on refining the model architecture, incorporating additional velocity constraints, and improving the diversity of the training simulations to enhance the model's accuracy and robustness.

Overall, the `NeHOD` framework represents a promising and exciting next step in the non-parametric modeling of galaxy-halo connections. Our framework can model individual satellites and effectively capture complex dependencies between halo and galaxy properties and the simulation parameters, while requiring minimal assumptions and scaling well with high-dimensional data. We envision various applications for `NeHOD`, including generating realistic mock catalogs for large-scale structure and strong lensing studies, cosmological parameter inference, and outlier detection. Additionally, `NeHOD` can be used to explore various DM models and astrophysical scenarios, offering valuable insights into the underlying physics of our universe.

Chapter 5

Synthetic Gaia surveys of Milky-Way-mass galaxies

Disclaimer

The content of this chapter was published in *The Astrophysical Journal* Volume 966, Number 1, Page 108 in April 2024 [174]. The authors are Tri Nguyen, Xiaowei Ou, Nondh Panithanpaisal, Nora Shipp, Lina Necib, Robyn Sanderson, and Andrew Wetzel. Authors Nguyen and Ou contribute equally to this work.

Abstract

The third data release (DR3) of *Gaia* has provided a five-fold increase in the number of radial velocity measurements of stars, as well as a stark improvement in parallax and proper motion measurements. To help with studies that seek to test models and interpret *Gaia* DR3, we present nine *Gaia* synthetic surveys, based on three solar positions in three Milky Way mass galaxies of the *Latte* suite of the FIRE-2 cosmological simulations. These synthetic surveys match the selection function, radial velocity measurements, and photometry of *Gaia* DR3, adapting the code base `Ananke`, previously used to match the *Gaia* DR2 release in Sanderson et al. 2020. The synthetic surveys are publicly available and can be found at http://ananke.hub.yt/. Similarly to the previous release of `Ananke`, these surveys are based on cosmological simulations and thus able to model non-equilibrium dynamical effects, making them a useful tool in testing and interpreting *Gaia* DR3.

The *Gaia* mission [412] has revolutionized the study of our Galaxy, the Milky Way. The second data release [DR2; 413] provided positions, proper motions, and parallaxes for

over one billion stars, including the first kinematic measurements of many stars across the Galaxy. In addition, DR2 included radial velocities for ~ 7 million stars, making it the largest six-dimensional kinematic catalog at the time. This data has enabled the discovery of new merging events, such as the *Gaia* Sausage Enceladus [69], [414], and the Kraken [415] (see [106] for a review), the construction of a new 3D dust map of the Milky Way [416], a detailed study of open clusters to unveil the history of the Galactic disk [417], an accurate measurement of the circular velocity of the Galaxy [418], and detailed studies of the fine resonances of the Milky Way disk [see e.g. 419]. The third data release [DR3; 420], [421] builds upon DR2, incorporating 12 months of additional observations, and significantly increasing the catalog of stars with 6D phase space measurements, including radial velocities, to ~ 33 million stars, as well as reducing uncertainties on and increasing the size of the sample of stars with full astrometry. These data have further enabled a deeper understanding of the dynamics of the Galaxy, for example by extending measurements of the circular velocity of the Milky Way to larger distances [52].

Synthetic catalogs and mock observations generated from cosmological simulations provide a valuable comparison to these rich observations of our own Galaxy. They enable tests of analysis tools and of our ability to recover true properties of our Galaxy from observations. [173], hereafter S20, produced nine *Gaia* DR2 synthetic surveys of the *Latte* suite of simulations [155], [422], using the code `Ananke`. Such synthetic surveys have been used in many studies involving the dynamics of the Milky Way, for example to estimate the detectability of simulated stellar streams [423], as a training set for a neural network that built the first accreted star catalog in the Milky Way [65], leading to a discovery of a prograde local structure Nyx [67], as a framework to test the ability of unsupervised machine learning techniques to reproduce the stellar phase space density [424], and as a link to connect the formation history and the components of the Milky Way [425]. In this work, we present synthetic *Gaia* DR3 surveys based on the same suite of *Latte* simulations.

The *Latte* simulations first introduced in [422] are baryonic zoom-in simulations of Milky Way analogs from the Feedback in Realistic Environments (FIRE) project [155], [426]. With an initial stellar particle mass resolution of $7070 M_\odot$, the *Latte* simulations resolve stellar populations down to the masses of individual star clusters. They self-consistently model baryonic processes, including star formation and the metal-enrichment of gas, which is essential for accurately calculating the extinction of observed stars. At the same time, they incorporate the effects of galaxy formation in a cosmological context, including a realistic history of mergers and accretion events.

`Ananke` is a framework for producing synthetic surveys based on the FIRE simulations, first presented by S20. Such work is based on Galaxia [427], which generated synthetic

surveys of the Milky Way based on kinematic distributions and N-body simulations. The framework entails sampling a population of individual stars from simulated star particles, assigning them realistic physical properties, and applying a simple error model to produce mock observations. `Ananke` has been applied to produce mock observations of other surveys from a range of simulated data sets, such as APOGEE [428], Dark Energy Survey [DES; 62], [429], and the Rubin Observatory Legacy Survey of Space and Time [LSST; 113] in [423].

In this chapter, we use `Ananke` to produce synthetic *Gaia* DR3 surveys of three Milky Way analogs from the *Latte* simulation suite, focusing on the updates to the surveys compared to S20. This chapter is organized as follows: In Section 5.1 we review the simulations and mock catalogs used in this work, in Section 5.2 we discuss the synthetic *Gaia* DR3 observations, and in Section 5.3, we present the resulting synthetic surveys, comparing them with those from *Gaia* DR2. We list the columns of the public release and their definitions in Section 5.3.3, and discuss the use cases and limitations of these synthetic surveys in Section 5.4.

5.1 Simulations and Mock Catalogs

In this section, we outline the different steps to build both a mock catalog and a synthetic survey. We first seek to define these two terms. Generating a *mock catalog* consists of spawning stars from star particles in the initial simulations. This process is independent of the target survey. The star particles in simulations typically have masses orders of magnitude greater than a single solar mass, depending on the initial simulation resolution. For the FIRE simulations, the star particles have a mass $\sim 7000\,M_\odot$. Generating a *synthetic survey* involves incorporating the specifics of a particular survey into the catalog of simulated stars, including, for example, photometric passbands, measurement errors, dust extinction, and the observer's location.

In order to build the new synthetic `Ananke` DR3 survey, we use the same three zoom-in simulations of Milky Way-mass galaxies from the *Latte* suite of FIRE-2 simulations as in S20 (m12i, m12f, m12m)[1]. The choice of these specific simulations is motivated by [422], [430], which have shown that these simulated galaxies Specifically, studies have shown that these simulated galaxies have galactic bar morphology [431], [432], stellar thin and thick disk morphology [433], gas kinematics [434], etc., that are broadly similar to the Milky Way. It is impossible, however, for any simulated galaxy to fully reproduce all properties of the Milky Way, and thus, we caution those using the synthetic surveys to be aware of the differences and refer to the literature mentioned above for details.

[1]For a brief overview of the simulation and the simulated galaxies used, we refer the reader to Section 2 of S20.

Table 5.1: The coordinates of each LSR as shown in Table 4 of S20.

label	$x_{\rm LSR}$ (kpc)	$y_{\rm LSR}$ (kpc)	$z_{\rm LSR}$ (kpc)	$v_{x,\rm LSR}$ (km/s)	$v_{y,\rm LSR}$ (km/s)	$v_{z,\rm LSR}$ (km/s)	$v_{R,\rm LSR}$ (km/s)	$v_{Z,\rm LSR}$ (km/s)	$v_{\phi,\rm LSR}$ (km/s)
`m12i-lsr-0`	0.0	8.2	0.0	224.7092	−20.3801	3.8954	−17.8	−3.9	224.4
`m12i-lsr-1`	−7.1014	−4.1	0.0	−80.4269	191.7240	1.5039	−24.4	−1.5	210.9
`m12i-lsr-2`	7.1014	−4.1	0.0	−87.2735	−186.8567	−9.4608	22.1	9.5	206.5
`m12f-lsr-0`	0.0	8.2	0.0	226.1849	14.3773	−4.8906	14.9	4.9	227.9
`m12f-lsr-1`	−7.1014	−4.1	0.0	−114.0351	208.7267	5.0635	−3.4	−5.1	244.3
`m12f-lsr-2`	7.1014	−4.1	0.0	−118.1430	−187.7631	−3.8905	−11.4	3.9	227.4
`m12m-lsr-0`	0.0	8.2	0.0	254.9187	16.7901	1.9648	16.2	−2.0	254.7
`m12m-lsr-1`	−7.1014	−4.1	0.0	−128.2480	221.1489	5.8506	2.4	−5.9	252.7
`m12m-lsr-2`	7.1014	−4.1	0.0	−106.6203	−232.2056	−6.4185	15.4	6.4	265.3

5.1.1 Locations of the Sun

To build a synthetic survey, we must assume the location of the observer, which we define as the solar position in the simulation. The procedure we adopt here for the coordinate transformation and the definition of the local standards of rest (LSRs) remain unchanged from S20, which we briefly summarize here. We assume that the Sun is at $R_\odot = 8.2$ kpc [435] in the three simulations[2], and define the principle axes based on the moment of inertia tensor of the youngest stars (with ages < 1 Gyr) located within $R_\odot$. We note that it is possible to scale the solar position differently for each simulation based on disk scale radii, local density, or the local circular velocity. As pointed out in S20, however, doing so introduces extra complexity for users to accommodate different solar radii for different simulations, and the variation in the results is not significant compared to changing the azimuth position of the Sun.

The three positions of the Sun in each galaxy are chosen to be evenly distributed in azimuthal angle to allow different view angles of axisymmetric features such as the bar and spiral arms, and at vertical distance $Z_\odot = 0$ kpc. We define the velocity of LSR as the median velocity of the star particles within 200 pc of the solar position. We summarize the positions and velocities of the LSRs in Table 5.1, which matches Table 4 of S20.

[2]This is an appropriate approximation given that these simulations have comparable scale heights and radii to the Milky Way.

5.1.2 Building a mock catalog

In this section, we discuss the procedure to build a *mock catalog* by converting the star *particles* from the FIRE-2 simulations into synthetic stars, leaving the construction of the *synthetic survey* in which we add the correct properties to these synthetic stars as drawn by the survey to Section 5.2.

Each star particle, with a mass $M_* \sim 7070\,\mathrm{M}_\odot$, is assumed to represent a population of synthetic stars with a single age and metallicity. To generate such mock catalogs, S20 used the GALAXIA code [427]. In this work, we adopt the same mock catalogs as in S20, modifying the stellar isochrones used in the generation of stars to the updated *Gaia* DR3 isochrones. This modification is performed at Step 2 below, while keeping the masses and the phase-space positions the same as in S20 in Step 1 & 3.

To build a mock catalog, we perform the following three steps. We will leave a detailed description of the DR3 isochrones to Section 5.2.1.

1. First, we sample the stellar masses of synthetic stars using the initial mass function (IMF) in [436] until the total mass equals the mass of the particle.[3]

2. Select the isochrone model that is closest in age and metallicity to the parent particle and obtain stellar properties and *Gaia* DR3 passband magnitudes by interpolating across initial stellar mass. Only stars with estimated unextincted apparent *Gaia* DR2/3 magnitudes of $3 < G < 21$ are kept in the catalog, before applying the full selection function in Section 5.2.4.

3. Assign the positions in phase space to each star by sampling from a parabolic or Epanechikov density kernel [437] centered on the parent particle. The density kernel generates smooth phase space distributions from the discrete individual parent particles. We then sample the position and velocities of the synthetic stars from the smooth distribution. The smoothing kernel is computed from the 6-dimensional phase space coordinates using the ENLINK code [438], [439]. Similarly to S20, we use the nearest 8 neighboring star particles to compute the kernel size. The kernel size taken with respect to two independent smoothing lengths, one for the distances, and one for the velocities. The size is proportional to the geometric mean of the smoothing lengths along each of the three-dimensions. To preserve the dynamic ranges of the different stellar populations and avoiding the oversmoothing of structures from different stellar

[3] The number of stars sampled is required to be an integer, while the fraction of the IMF within a subrange of mass is not. Some rounding is assumed, and given that the highest possible stellar mass is still two orders of magnitude lower than the mass of the star particle, in this work, as in S20, we assume that this is a valid approach with a small fractional error.

populations, a kernel is computed for *in situ* stars, which are defined as those formed within 30 physical kpc of the main galaxy, while a separate kernel is computed for stars formed outside of this radius. In addition, we subdivide *in situ* stars into eight age bins corresponding to the populations of the Besançon Milky Way model (Table 2.1 of [440]) and compute a different kernel for each of them. For full details on the phase space smoothing procedure, we refer readers to Section 4.3 of S20. This kernel is not optimized for small-scale structures, and in some cases may introduce unphysical features into substructures such as low mass satellite galaxies and stellar streams. For example, [423] adopted a different kernel (albeit also based on the Epanechikov kernel) but with the 16 neighboring star particles and a kernel size that is inversely proportional to the cube-root of the local density around each parent particle, to properly smooth out stellar streams.

5.2 Synthetic Surveys

We describe the procedure used to produce the `Ananke` DR3 *synthetic surveys*. As mentioned in Section 5.1, we use the mock catalogs presented in S20, and apply updated DR3 isochrones (Section 5.2.1), extinction modeling (Section 5.2.2), observational uncertainty modeling (Section 5.2.3), and selection function (Section 5.2.4).

5.2.1 Isochrones

We use updated *Gaia* DR3 passbands and isochrones from Padova CMD v3.6[4] to generate updated intrinsic *Gaia* DR3 magnitudes for stars in the mock catalogs in the G, G_{BP}, and G_{RP} bands. The photometric system follows the revised and expanded library described in [441], adopting a revised spectral energy distribution (SED) for Vega from [442]. Two assumptions are made while adopting the isochrones. First, circumstellar dust is ignored as it mostly affects the bright end of the isochrones, where the grid is the sparsest. Therefore, linear interpolation with the circumstellar dust included creates unphysical features when stars fall between these sparse grid points. Second, we remove the isochrone grid points representing white dwarfs, as the transition from the tip of the red giant branch to the white dwarf is not modeled by Galaxia. Since Galaxia takes the edge value for magnitudes when a star is outside of the isochrone grid, stars beyond the last non-white dwarf grid point are all assigned the same magnitudes, creating artificial overdensities at the tip of the giant branches in the final sample. We expect these stars to be potential white dwarfs and flag

[4] http://stev.oapd.inaf.it/cgi-bin/cmd

all affected stars in the final synthetic survey as `flag_wd` and recommend removing stars with `flag_wd` set to 1 before conducting analysis. We expect a minimal effect on the overall completeness of the sample as a result of this treatment. As shown by [443], 359,073 white dwarfs are confidently identified in *Gaia* DR3, comprising less than 0.05% of the full *Gaia* sample. Even accounting for the fact that the white dwarf catalog presented by [443] is less complete in crowded regions near the galactic plane, the total white dwarf count in the actual *Gaia* catalog is expected to be a tiny fraction of the full sample. Thus, for the synthetic survey, the overall loss in stellar count and impact on sample completeness are expected to be minimal as a result of this cut.

5.2.2 Extinction Modeling

We adopt a self-consistent extinction model similar to Section 5.1 of S20, which we briefly describe below. The FIRE-2 simulations do not resolve the creation and destruction of dust grains, so we assume the line-of-sight extinction by dust traces the metal-enriched gas in the simulations. We calculate the reddening $B-V$ of each star using the metal-weighted column density of hydrogen along the line of sight between the star and the solar position. The extinction is therefore calculated self-consistently, using the gas and metal distributions of each individual simulated galaxy, and thus accurately captures the spatial structures of the galaxy (see Figure 12 of S20 for the simulated galaxy dust map). The extinction at 550 nm A_0 is calculated using the standard relation, $A_0 = 3.1E(B-V)$ [444]–[446], and then converted into extinction in the *Gaia* DR3 passbands.

Using the coefficient A_0 from the `Ananke` DR2 mock catalogs, as described in [173], we re-calculate the extinction coefficients $A_{G,BP,RP}$ in the *Gaia* DR3 passbands. We adopt the extinction conversion relation provided by the *Gaia* collaboration as part of the auxiliary data for eDR3 to compute $A_{G,RP,BP}$ as functions of A_0 and the unextincted color $(G_{BP}-G_{RP})$. [5]. Specifically, we compute $A_{G,BP,RP} = k_{G,BP,RP}\,A_0$, where $k_{G,BP,RP}$ is a function of A_0 and $(G_{BP}-G_{RP})$. Using the extinction coefficients $(A_{G,BP,RP})$, we convert the intrinsic magnitudes interpolated from the isochrones into the extincted intrinsic magnitudes. These extincted intrinsic magnitudes are combined with the distance modulus to calculate the true extincted apparent magnitudes.

Following the recommendation from the *Gaia* collaboration, we do not directly apply the extinction law outside of the applicable color range, $-0.06 < G_{\rm BP} - G_{\rm RP} < 2.5$. However, excluding stars outside of range introduces an unnatural cut on the $G_{\rm BP}-G_{\rm RP}$ distribution.

[5]The relationship and coefficients can be downloaded from https://www.cosmos.esa.int/web/gaia/edr3-extinction-law

We therefore extrapolate their *Gaia* passbands extinction coefficients using the nearest $G_{\rm BP} - G_{\rm RP}$ extreme value (i.e. -0.06 or 2.5).

The extinction law is also limited to extinction coefficients (A_0) in the range from 0.01 to 20. On the low end, since the extinction law returns finite positive values for $k_{G,BP,RP}$, the resulting $A_{G,BP,RP}$ always converge to 0 as A_0 goes to 0 and thus the extinction law naturally extrapolates to $A_0 = 0$. On the high end, unlike with the color, we do not adopt the extreme value (i.e., 20) for A_0 or attempt to approximate the extinction law outside of the applicable A_0 range. Stars with $A_0 > 20$ are not included in the final synthetic survey for two reasons. Firstly, A_0 is implicitly related to the distance of the star as the extinction arises from the dust between the star and the observer. If we were to adopt the extreme value for the extinction coefficient for a given star, the reported extincted photometry in the final synthetic survey would be inconsistent with the reported parallax and the underlying dust map. Secondly, the extincted apparent magnitudes for the majority of the stars with $A_0 > 20$ are expected to be significantly fainter than the observational limit of the synthetic survey ($G_{\rm obs} < 21$, as described in Section 5.2.4). We therefore do not expect the cut on $A_0 > 20$ to have a significant impact on the completeness of the final synthetic survey.

5.2.3 Error Modeling

We construct the photometric error model from the fit *Gaia* DR3 photometric uncertainties tool provided by *Gaia* DPAC (Data Processing and Analysis Consortium)[6], based on data originally described in [184]. We adopt the astrometric measurement error models from the `PyGaia` package[7]. The spectroscopic error model is obtained from private communication with the *Gaia* collaboration as a function of $T_{\rm eff}$ and $G_{\rm RVS}$. We calculate the errors and the error-convolved quantities by randomly sampling from a one-dimensional Gaussian centered on the truth values. In the final catalog, we report both the truth values and the error-convolved values.

Photometric Error

As mentioned, we adopt the photometric uncertainties tool from *Gaia* DPAC to calculate the errors in (G, G_{BP}, G_{RP}). The tool models the median behavior of the real *Gaia* (e)DR3 photometric uncertainties in the three *Gaia* passbands via cubic B-spline fitting. The errors in each photometric band are calculated as a function of the band *extincted* magnitudes. Because the B-spline is restricted to a range $[4, 21]$ in all three bands, we extrapolate the

[6] https://www.cosmos.esa.int/web/gaia/fitted-dr3-photometric-uncertainties-tool
[7] https://github.com/agabrown/PyGaia

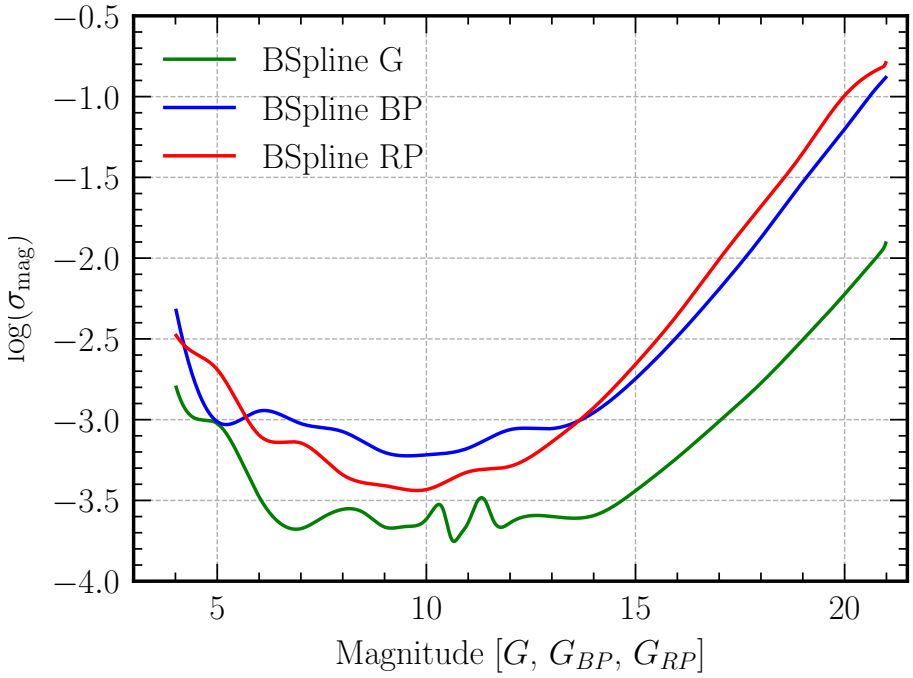

Figure 5.1: Logarithm of the expected uncertainties for sources with 200(20) observations in $G(G_{BP}/G_{RP})$.

Table 5.2: Coefficients of the astrometric errors in Equation 5.1.

ϖ	$\alpha_\star$	δ	$\mu_{\alpha\star}$	μ_δ
1.0	0.80	0.70	1.03	0.89

photometric uncertainties of each band using the nearest extreme values (i.e. 4 or 21) In addition, the tool is capable of scaling the fit B-splines with different numbers of observations. We take, for simplicity, the default number of observations (i.e., 200 for G and 20 for G_{BP}/G_{RP}) for all stars in our catalogs. We show in Figure 5.1 the errors as a function of extincted magnitude and reproduce Figure 14 of [184]. We note that this error modeling does not take into account systematic effects originating from the properties of the source, e.g., position and color.

Astrometric Error

`PyGaia` models the astrometric errors (i.e. parallax, position, and proper motion) as solely dependent on the apparent G magnitude. The position and proper motion errors are returned in the ICRS frame, i.e. in RA and Dec. To obtain the error-convolved positions and proper

Table 5.3: Coefficients for color transformation from $G - G_{RP}$ to $G_{RVS} - G_{RP}$.

a_0	a_1	a_2	a_3	$G - G_{\rm RP}$ range
−0.0397	−0.2852	−0.0330	−0.0867	$[-0.15, 1.2]$
−4.0618	10.0187	−9.0532	2.6089	$[1.2, 1.7]$

Table 5.4: Coefficients for $\sigma_{\rm RV}$ as a function of $G_{\rm RVS}$

$\sigma_{\rm floor}$	a	b	$G_{\rm RVS,0}$	Applicable range
0.12	0.9	6.0	14.0	$T_{\rm eff} < 6500$ K
0.4	0.8	20.0	12.75	$T_{\rm eff} > 7000$ K

motions in Galactic coordinates (ℓ, b), we calculate the error-convolved ICRS coordinates and apply a coordinate transformation. Similarly to *Gaia*, we do not report the error in the Galactic coordinates. The astrometric errors, $\sigma_{\rm X}$, for $X \in (\alpha_\star, \delta, \mu_{\alpha\star}, \mu_\delta)$, can be summarized as follows:

$$\sigma_X = c_X \sigma_\varpi, \quad \sigma_\varpi = \sqrt{40 + 800z + 30z^2} \tag{5.1}$$

$$\log_{10} z = 0.4(\max[G, G_{\rm bright}] - 15.0), \tag{5.2}$$

where σ_ϖ is the parallax error, and $G_{\rm bright} = 13$. The coefficients c_X are reported in Table 5.2. Because `PyGaia` returns the error in RA cos(Dec) $\sigma_{\alpha\star}$, we convert $\sigma_{\alpha\star}$ to the RA error σ_α via analytical error propagation.

Spectroscopic Error

For spectroscopic measurements, *Gaia* DR3 provides radial velocity spectra (with magnitude $G_{\rm RVS}$), object classifications, and measured stellar parameters, such as effective temperature, surface gravity, extinction coefficient, and metallicity, in addition to radial velocities. Our synthetic survey only provides error-convolved radial velocity measurements. For DR3 radial velocities, we first use relationships provided by [181] to obtain true *Gaia* RVS magnitude, G_{RVS}, from G and G_{RP}. To do so, we use

$$\begin{aligned} G_{RVS} - G_{RP} = a_0 &+ a_1(G - G_{RP}) \\ &+ a_2(G - G_{RP})^2 + a_3(G - G_{RP})^3, \end{aligned} \tag{5.3}$$

where the coefficients are provided in Table 5.3. As for the extinction law extrapolation, we approximate the conversion for stars outside of the applicable range ($-0.15 < G - G_{\rm RP} < 1.7$), using the coefficients corresponding to the nearest $G - G_{\rm RP}$ extreme value (i.e., -0.15 or

Table 5.5: Coefficients for f as a function of $G_{\rm RVS}$

a	b	c	Applicable range
0.318	0.3884	-0.02778	$G_{\rm RVS} < 12$ mag
16.554	-2.4899	0.09933	$G_{\rm RVS} > 12$ mag

1.7).

The radial velocity uncertainty is fit as a function of $G_{\rm RVS}$,

$$\sigma_{\rm RV} = \sigma_{\rm floor} + b \exp\left(a(G_{\rm RVS} - G_{\rm RVS,0})\right). \tag{5.4}$$

The coefficients a, b are fit independently for cooler ($T_{\rm eff} < 6750$ K) and warmer ($T_{\rm eff} > 6750$ K) stars (Table 5.4), obtained from private communication with the Gaia Collaboration prior to the official release of the third data release. Warmer stars generally have a larger error in radial velocities. While the error modeling for warm stars, as shown in Figure 5.2, appears to greatly exceed 10 km s^{-1}at the very faint end ($G_{\rm RVS} \sim 14$), we note that only warm stars with $G_{\rm RVS} < 12$ are selected to have a measured radial velocity in the final catalog, as described in more detail in Section 5.2.4. The maximum radial velocity measurement uncertainties are thus $\sim$6 km/s for cool stars and $\sim$11 km/s for warm stars.

During the preparation of this manuscript, we were made aware of a more detailed *Gaia* DR3 radial velocity error model based on the derived stellar population.[8] Since our synthetic survey does not include the stellar evolutionary stage, we opt for the simple recipe that assigns errors based on the effective temperature of the stars. The coefficients provided are not identical to those adopted here, but a comparison between the two indicates that our adopted error modeling is roughly consistent with those from the more detailed model. For the most part, our error modeling falls on the conservative side of the latest model.

[182] noted that during scientific validation of the published DR3 radial velocities, the above uncertainties were underestimated and thus require an additional multiplicative correction factor f. This multiplicative factor (f) is a function of $G_{\rm RVS}$,

$$f = a + b\, G_{\rm RVS} + c\, G_{\rm RVS}^2, \tag{5.5}$$

with coefficients given in Table 5.5. The velocity uncertainties should therefore be $f \times \sigma_{V_{\rm R}}$. We note that the relation is only valid for $G_{\rm RVS} > 8$. For $G_{\rm RVS} < 8$, we still apply the correction function but assume $G_{\rm RVS} = 8$. The correction factor is not applied directly to the uncertainties in the final *Gaia* DR3 dataset. Following that practice, we calculate this

[8] https://www.cosmos.esa.int/web/gaia/science-performance

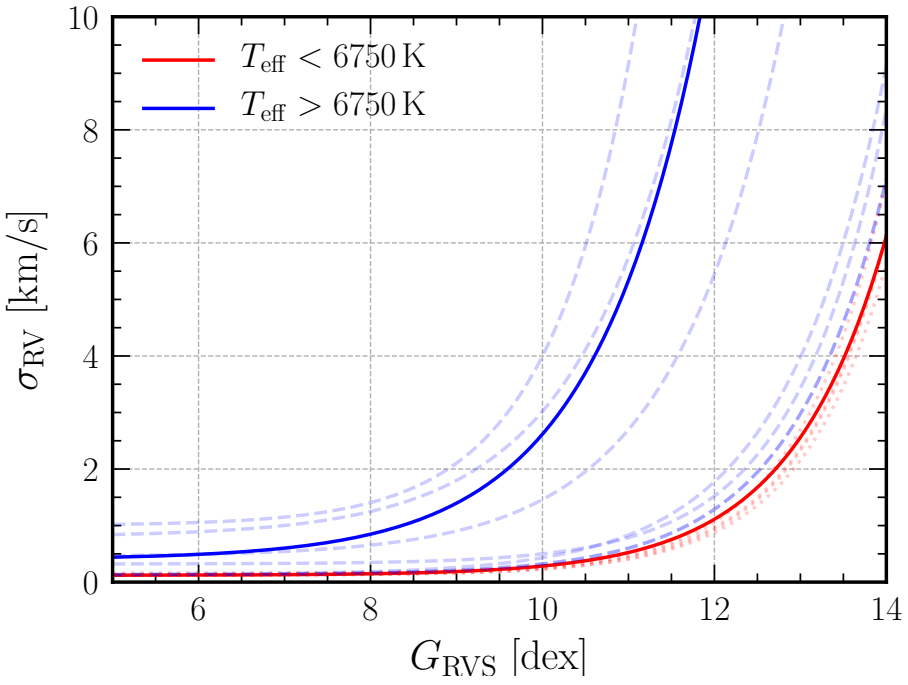

Figure 5.2: Comparison between the radial velocity error models adopted in this study (solid line) and those provided by the *Gaia* collaboration with the official release of DR3 (dashed/dotted lines). Blue dashed lines represent error estimates for dwarfs, whereas red dotted lines represent error estimates for giants. Our estimate for cooler stars ($T_{\rm eff} < 6750$ K) is largely consistent with the estimates for giants and, similarly, our estimate for warmer stars ($T_{\rm eff} > 6750$ K) with the estimates for dwarfs.

correction factor and provide it separately in the final synthetic survey.

5.2.4 Selection Function and Data Release

With error-convolved values computed, we next apply the selection function to produce the final synthetic surveys. We apply two selection functions, one for selecting stars that are detectable in all three photometric bands and another for selecting stars with reported radial velocity.

We apply a G-band magnitude cut to select stars with reported photometry in each catalog. We note that the cuts are applied on the error-convolved observed magnitudes. We select the sample of stars with reported photometry via a cut on the observed G magnitude, $3 < G_{\rm obs} < 21$. This is the same selection cut applied in S20.

To select the sample of stars with reported radial velocities, we make a cut on effective temperature, $T_{\rm eff}$ and $G_{\rm RVS}$. S20 reported radial velocity measurements for bright stars with $G_{\rm RVS} < 14$ and effective temperature of $3550 < T_{\rm eff} < 6900$ K. We extend the radial velocity selection to $3600 < T_{\rm eff} < 14500$ K for bright stars ($G_{\rm RVS} \leq 12$) and $3100 < T_{\rm eff} < 6750$ K for fainter stars ($12 < G_{\rm RVS} < 14$), in order to match the temperature range reported in [182], reflecting the improvements from *Gaia* DR2 to DR3.

We bin the stars in each catalog by their LSR-centric distance into 10 radial slices. Table 5.6 shows the total number of stars, as well as the number of stars with radial velocity measurements, in each radial slice and catalog.

Table 5.6: Number of stars in the `Ananke` DR3 surveys of the *Latte* MW-mass suite of FIRE simulations.

File Information			Number of stars					
	$d_{\rm min}$	$d_{\rm max}$	m12i			m12i with radial velocity		
index	[kpc]	[kpc]	lsr-0	lsr-1	lsr-2	lsr-0	lsr-1	lsr-2
0	0	3	316,095,707	357,883,822	392,451,308	15,646,316	18,615,906	20,116,160
1	3	4.25	290,904,524	299,505,191	302,616,084	5,468,516	5,571,554	5,555,894
2	4.25	5.5	401,479,587	420,118,792	371,841,356	7,015,408	7,442,827	6,477,804
3	5.5	6.5	400,845,878	451,022,716	362,386,254	6,624,552	7,557,280	6,134,119
4	6.5	7.25	365,130,175	436,233,406	320,076,487	6,257,625	7,157,334	5,952,305
5	7.25	8	418,818,886	526,227,703	362,084,404	8,324,411	9,089,852	8,126,577
6	8	9	507,799,164	823,576,930	435,396,583	11,999,243	14,032,870	11,840,268
7	9	10	320,749,442	382,157,501	269,564,468	8,105,168	8,450,961	7,978,302
8	10	15	510,906,338	604,188,984	455,152,086	15,677,739	15,913,215	15,528,607
9	15	300	149,436,821	137,907,134	133,845,033	11,557,428	9,579,237	10,301,140
		Total	3,682,166,522	4,438,822,179	3,405,414,063	96,676,406	103,411,036	98,011,176
		DR2 Total	3,215,565,725	3,754,501,977	2,932,162,112	38,183,839	44,583,007	39,191,496

Table 5.6: (continued)

File Information			Number of stars					
	$d_{\rm min}$	$d_{\rm max}$	m12f			m12f with radial velocity		
index	[kpc]	[kpc]	lsr-0	lsr-1	lsr-2	lsr-0	lsr-1	lsr-2
0	0	3	424,006,559	608,837,889	547,265,888	20,533,729	31,377,347	26,192,624
1	3	4.25	399,657,145	422,500,160	434,830,068	7,581,674	7,891,448	8,082,358
2	4.25	5.5	577,804,699	526,740,558	506,589,056	11,046,401	9,879,024	9,407,981
3	5.5	6.5	606,129,500	519,988,527	484,524,451	11,670,711	9,207,476	8,932,140
4	6.5	7.25	590,389,532	475,891,126	438,928,174	11,559,183	9,064,200	8,673,940
5	7.25	8	719,850,219	590,388,390	532,925,594	14,469,396	13,004,977	11,436,403
6	8	9	883,408,757	747,272,883	671,353,993	18,926,746	19,388,311	15,961,569
7	9	10	554,034,772	416,802,559	401,635,820	11,909,320	11,913,256	10,193,449
8	10	15	1,099,412,587	719,518,336	772,769,865	21,025,371	20,362,945	17,498,773
9	15	300	536,944,872	389,292,686	500,404,364	18,697,669	16,215,180	14,172,354
		Total	6,391,638,642	5,417,233,114	5,291,227,273	147,420,200	148,304,164	130,551,591
		DR2 Total	5,851,407,276	4,706,540,756	4,678,842,172	62,673,864	61,393,185	57,808,862

Table 5.6: (continued)

File Information			Number of stars					
	d_{min}	d_{max}	m12m			m12m with radial velocity		
index	[kpc]	[kpc]	lsr-0	lsr-1	lsr-2	lsr-0	lsr-1	lsr-2
0	0	3	984,809,951	1,073,978,992	910,734,608	47,393,328	54,240,600	43,119,764
1	3	4.25	728,265,777	798,150,011	686,462,276	12,171,491	13,592,305	11,429,030
2	4.25	5.5	814,806,044	863,540,944	796,780,191	12,767,460	13,727,680	12,465,994
3	5.5	6.5	685,954,361	723,050,215	689,642,062	10,229,572	10,706,710	9,751,780
4	6.5	7.25	528,436,556	558,951,415	531,816,221	7,774,165	7,855,599	7,300,150
5	7.25	8	523,399,484	551,230,847	532,527,598	7,611,444	7,589,807	7,471,248
6	8	9	2,003,093,353	639,727,826	617,194,075	17,707,039	9,881,911	10,009,474
7	9	10	422,716,282	458,827,031	432,726,049	8,073,088	8,469,156	8,438,569
8	10	15	835,507,954	1,267,343,926	1,192,679,933	23,167,408	26,128,219	26,126,666
9	15	300	261,056,409	268,075,320	244,124,886	20,074,951	20,492,819	19,063,333
		Total	7,788,046,171	7,202,876,527	6,634,687,899	166,969,946	172,684,806	155,176,008
		DR2 Total	5,701,759,381	6,415,674,623	5,516,835,110	84,931,532	108,808,464	78,520,886

5.3 Results

5.3.1 Comparison with Ananke DR2

We compare our final synthetic survey for *Gaia* DR3 using FIRE with the synthetic *Gaia* DR2 survey from S20. We updated the photometry to be consistent with *Gaia* DR3, using isochrones and extinction laws corresponding to the *Gaia* DR3 photometric system. We also updated the error modeling for photometric measurements and radial velocity measurements.

The detailed numbers of sources in each radial bin of each galaxy are given in Table 5.6. In general, there is a small increase in the total number of observed stars in the DR3 catalogs as compared to the DR2 catalogs. The number of stars with radial velocity measurements in each catalog has increased by $\sim 2-3$ times, as expected from the wider range of effective temperature $T_{\rm eff}$ in the selection cut (see Section 5.2.4). However, this is a more moderate increase than the factor of 5 between the two *Gaia* data releases (from ~ 7 millions stars in *Gaia* DR2 to ~ 33 million stars in *Gaia* DR3) [182]. This is due to the radial velocity selection cut in S20 being overly optimistic, already at $G_{\rm RVS} < 14$ when considering the actual performance of *Gaia* DR2 at $G_{\rm RVS} < 12.5$ [447]. In all catalogs, the overall fraction of stars with radial velocity measurement compared to the total sample, which can be calculated from Table 5.6, is about $2-3\%$, which is indeed comparable to that of *Gaia* DR3, which was about $\sim 2\%$ [182]. For reference, the fractions of stars with radial velocity measurements in the `Ananke` DR2 catalogs are about $1-1.5\%$.

5.3.2 Synthetic Surveys

In Figure 5.3 and Figure 5.4, we compare the distributions of radial velocity errors and proper motion errors between DR2 and DR3 for all stars in `m12i-lsr0-rslice0`[9]. Figure 5.3 shows the distributions of the radial velocity errors for DR2 (blue) and DR3 (solid black) for all stars with radial velocities. As expected, the radial velocity errors in `Ananke` DR3 are significantly lower than in DR2. The radial velocity errors for DR3 are composed of two stellar populations: one with low $T_{\rm eff}$ and one with high $T_{\rm eff}$, while the radial velocity errors in DR2 are modeled by a single exponential [173]. In `Ananke` DR3, the low $T_{\rm eff}$ population makes up most of the distribution below $\sigma_{V_{\rm R}} \lesssim 6$ km/s, while the high $T_{\rm eff}$ population is responsible for the tail at high $\sigma_{V_{\rm R}} \gtrsim 6$ km/s. The sharp cut at the lower end of the DR2 error is the systematic noise floor at 0.11 km/s mentioned in S20. Figure 5.4 shows the distributions of errors in the proper motions $\mu_{\alpha,\star}$ and μ_δ for DR2 (blue) and DR3 (solid

[9]The choice of the m12i and lsr-0 synthetic survey is just an example that we use to illustrate different properties. Similar treatment can be done with any of the other synthetic survey.

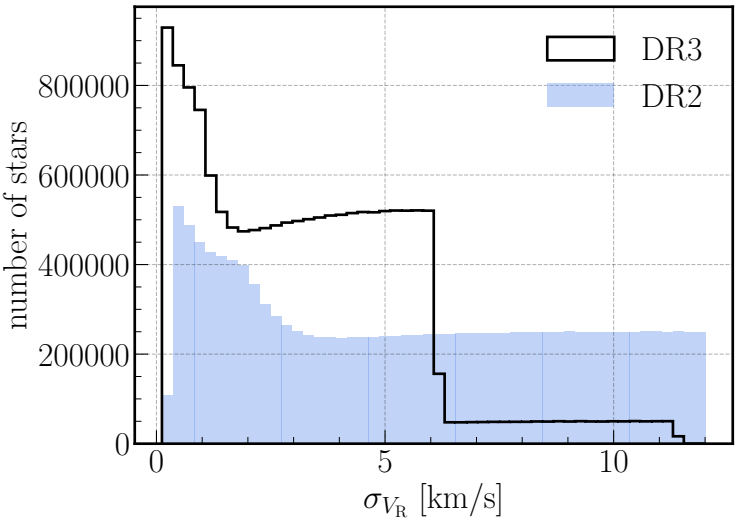

Figure 5.3: Distributions of the radial velocity errors for DR3 (solid black) and DR2 (blue) for all stars with radial velocities in `m12i-lsr0-rslice0`.

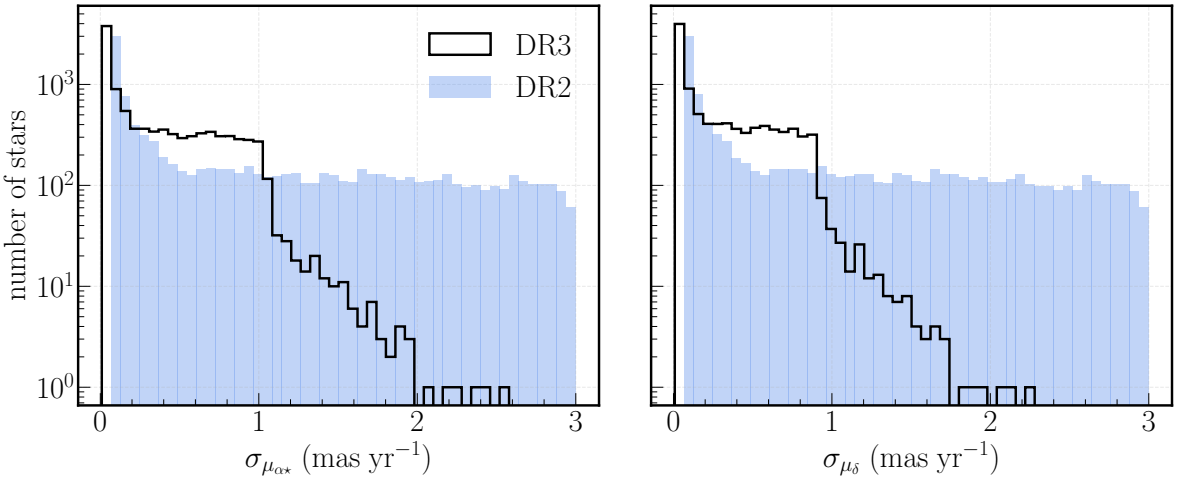

Figure 5.4: Distributions of the proper motion errors for RA cos(Dec) $\alpha^\star$ (left) and Dec δ (right) for DR3 (solid black) and DR2 (blue) for all stars in `m12i-lsr0-rslice0`.

black). Similarly, as with the radial velocity errors, the proper motion errors in `Ananke` DR3, shown in Fig. 5.4 are typically much lower than in DR2. The DR2 proper motion errors have a cut off of 0.0861852 mas/yr at the low end, as described in Equation 16 and Table 5 of S20.

We examine the Hertzsprung-Russell diagram of one simulation (m12i at rslice 0) as shown in Figure 5.5. We plot only stars with an estimated parallax error of less than 10%. Figure 5.5 shows that our results are qualitatively similar to what was shown in `Ananke` DR2 from S20. We see minor differences at the low-mass end of the main sequence and at the tip of the red giant branch, likely arising from the updated isochrones. Some echoes of the underlying grid of isochrones are still visible at the brightest magnitudes, where the model grid is sparsest, and potential artifacts from linear isochrone interpolation near the tip of the red giant branch are present (see Section 5.2). The diagrams for the other simulations presented in this study are shown in Figure 5.6.

We additionally compare our results with an actual *Gaia* DR3 CMD from [183]. In Figure 5.7, we partially reproduce Figure 1 in [183]. Our synthetic survey generates CMDs qualitatively similar to the *Gaia* DR3 data. When we only consider the subsample with RV measurements (`has_rvs`), the synthetic survey distributions qualitatively resemble that of the real *Gaia* DR3 survey.

5.3.3 List of parameters in the synthetic surveys

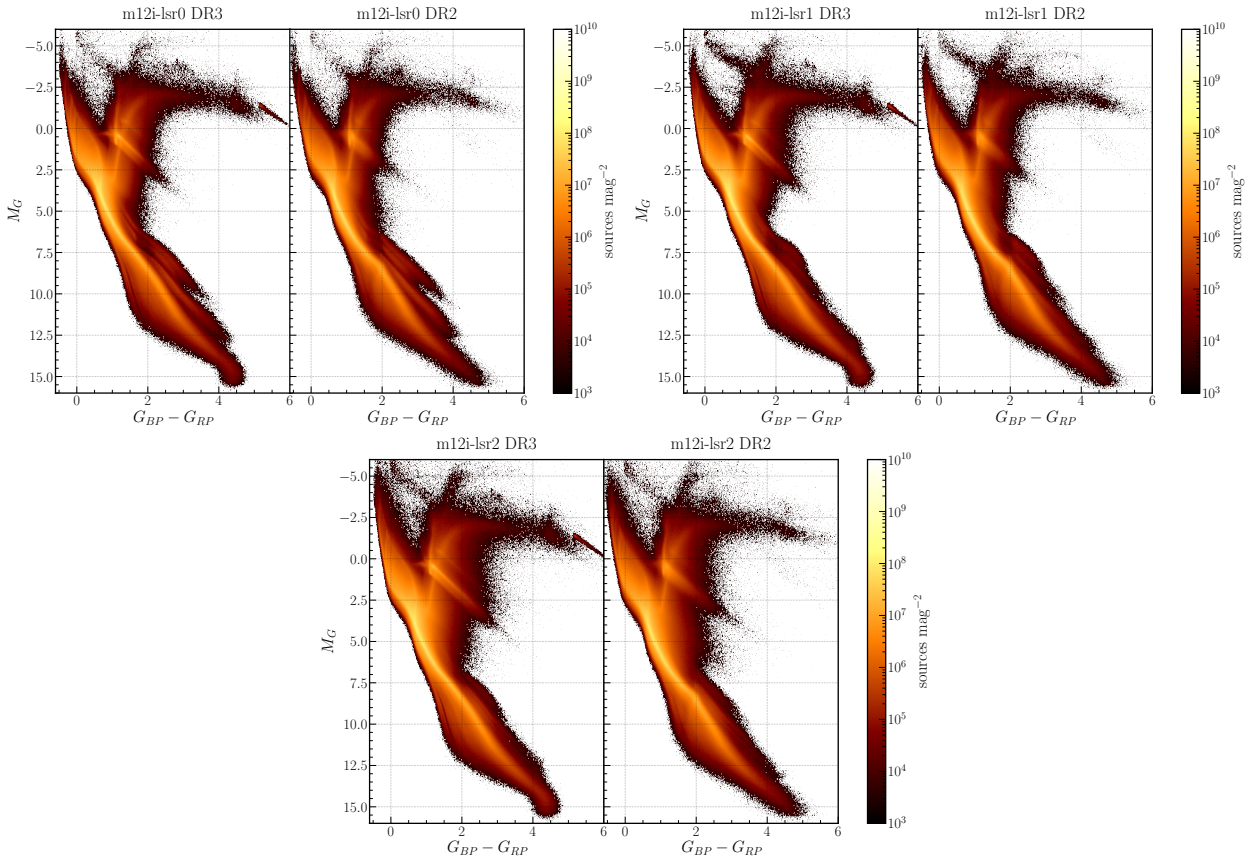

Figure 5.5: The Hertzsprung–Russell diagram for the first radial slice `rslice-0` of `m12i-lsr0`, `m12i-lsr1`, and `m12i-lsr2` for stars satisfying the parallax cut $\sigma_\varpi/\varpi > 10$.

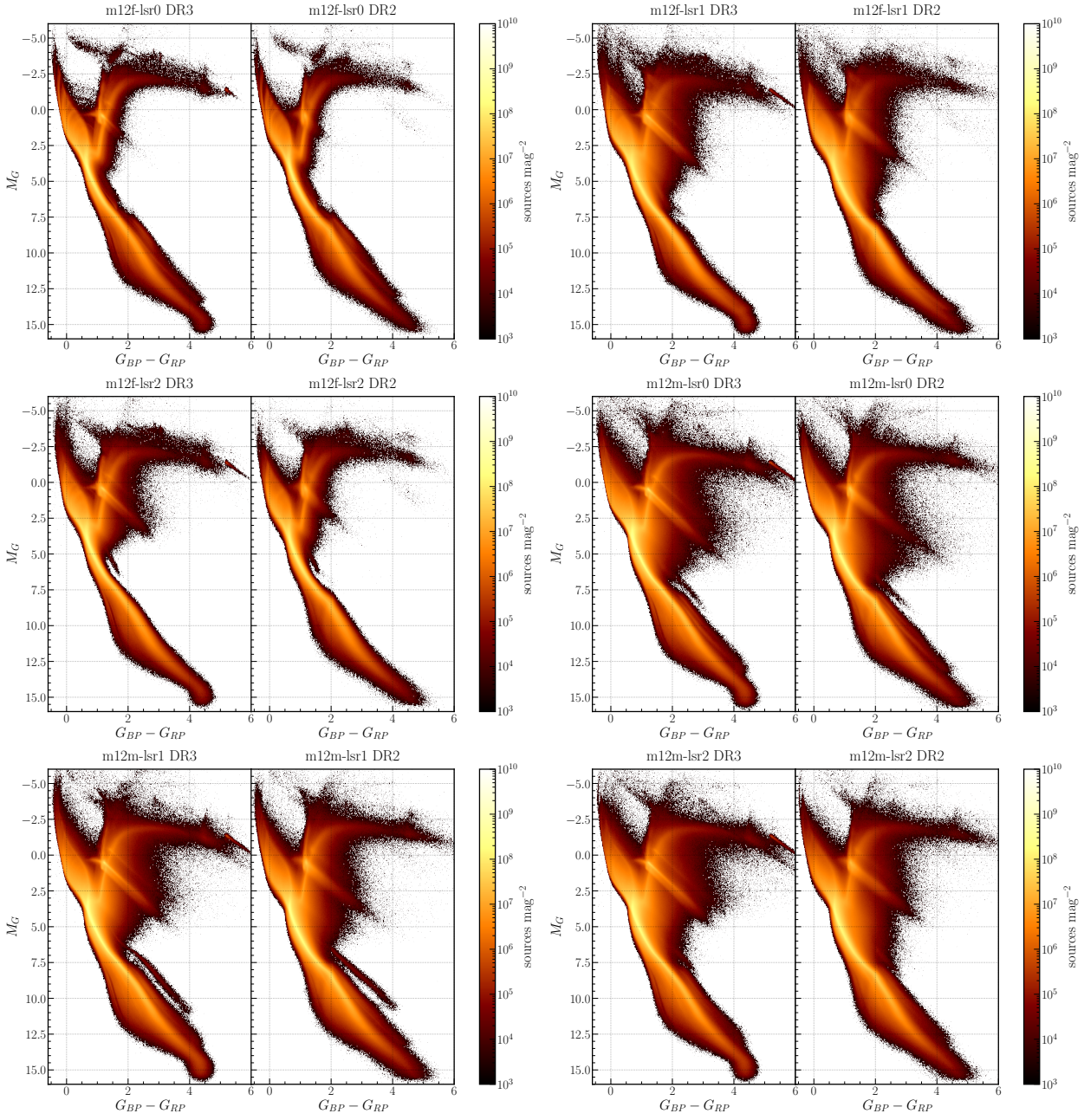

Figure 5.6: Same as Figure 5.5 but for `rslice-0` of `m12f` and `m12m` for all three LSRs.

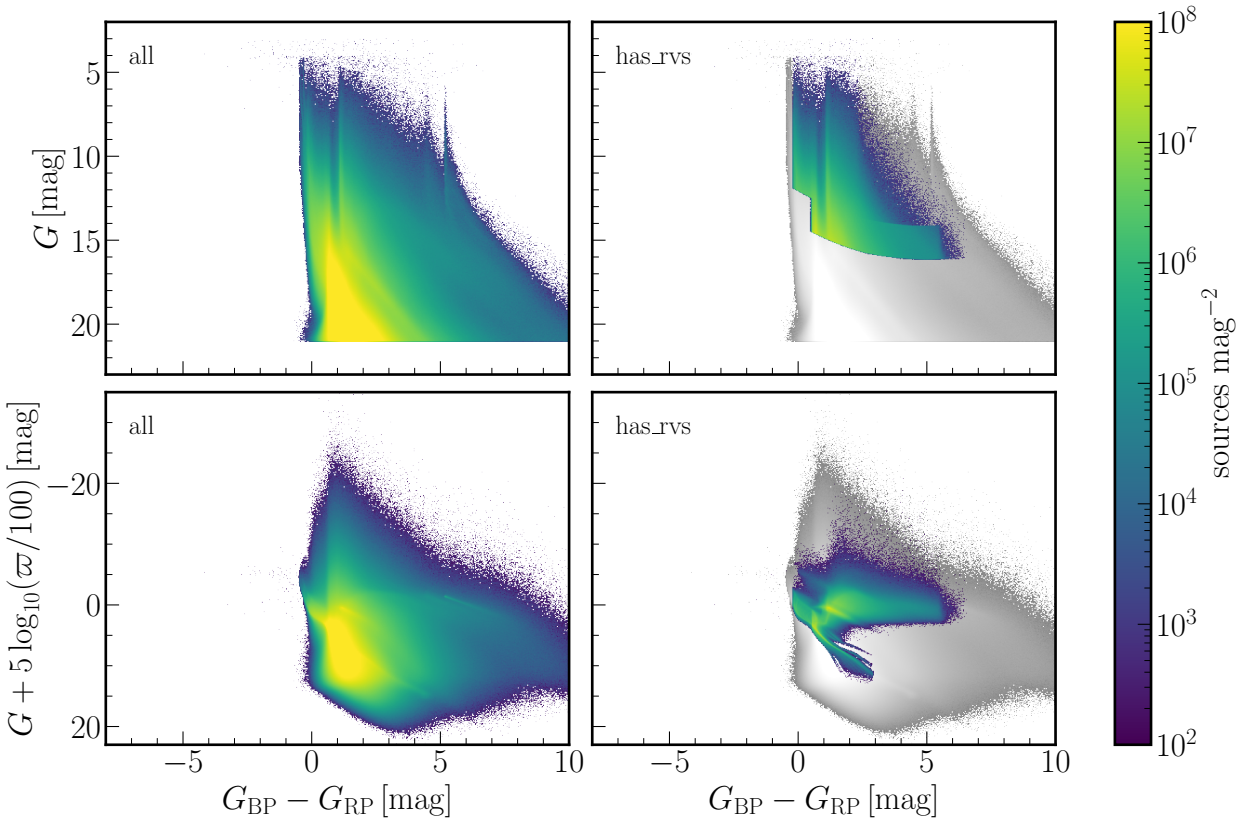

Figure 5.7: The color-magnitude diagram for `m12i-lsr0-rslice0` to `m12i-lsr0-rslice9` for all stars (left) and those with RV measurements (right) satisfying the positive parallax cut $\varpi > 0$. The top panels show the observed color-magnitude diagram, whereas the bottom panels show the absolute G magnitude computed from the measured parallax.

Table 5.7: Data Model for Synthetic Surveys.

Quantity	Explanation	Data type	Unit
	Fields with names identical to those in DR2		
	Astrometry		
`ra`	Right ascension	double	Angle (deg)
`ra_error`	Standard error of R.A.	double	Angle (deg)
`dec`	Declination	double	Angle (deg)
`dec_error`	Standard error of declination	double	Angle (deg)
`parallax`	Parallax	double	Angle (mas)
`parallax_error`	Standard error of parallax	double	Angle (mas)
`parallax_over_error`	Parallax divided by its error	float	
`pmra`	Proper motion in R.A. direction	double	Angular Velocity (mas yr^{-1})
`pmra_error`	Standard error of proper motion in R.A. direction	double	Angular Velocity (mas yr^{-1})
`pmdec`	Proper motion in Dec direction	double	Angular Velocity (mas yr^{-1})
`pmdec_error`	Standard error of proper motion in Dec direction	double	Angular Velocity (mas yr^{-1})
`l`	Galactic longitude (converted from R.A., Dec)	double	Angle (deg)
`b`	Galactic latitude (converted from R.A., Dec)	double	Angle (deg)
	Photometry		
`phot_g_mean_mag`	Extincted apparent G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag`	Extincted apparent G_{BP}-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag`	Extincted apparent G_{RP}-band mean magnitude	float	Magnitude (mag)
`bp_rp`	Reddened $G_{BP} - G_{RP}$ color	float	Magnitude (mag)
`bp_g`	Reddened $G_{BP} - G$ color	float	Magnitude (mag)
`g_rp`	Reddened $G - G_{RP}$ color	float	Magnitude (mag)

Table 5.7: (continued)

Quantity	Explanation	Data type	Unit
	Spectroscopy		
radial_velocity	Radial velocity	double	Velocity (km s^{-1})
radial_velocity_error	Standard error of radial velocity	double	Velocity (km s^{-1})
	Other fields not in the Gaia DR2 data model		
	Indices		
starid	array index of the star (per mock catalog)	long	
parentid	array index of the generating star particle in the snapshot file	long	
partid	0 if phase-space coordinates are identical to the generating star particle, 1 otherwise	short	
	Phase Space		
ra_true	true R.A.	double	Angle (deg)
dec_true	true Dec	double	Angle (deg)
dmod_true	true distance modulus	double	Magnitude (mag)
ra_cosdec_error	standard error in R.A.cos (Dec)	double	Magnitude (deg)
parallax_true	true parallax	double	Angle (mas)
pmra_true	true pm in R.A. direction	double	Angular Velocity (mas yr^{-1})
pmdec_true	true pm in Dec direction	double	Angular Velocity (mas yr^{-1})
radial_velocity_true	true RV	double	Velocity (km s^{-1})
l_true	true Galactic long	double	Angle (deg)
b_true	true Galactic lat	double	Angle (deg)
pml	pm in Galactic long direction	double	Angular Velocity (mas yr^{-1})
pmb	pm in Galactic lat direction	double	Angular Velocity (mas yr^{-1})
pml_true	true pm in Galactic long direction	double	Angular Velocity (mas yr^{-1})
pmb_true	true pm in Galactic lat direction	double	Angular Velocity (mas yr^{-1})

Table 5.7: (continued)

Quantity	Explanation	Data type	Unit
`px_true, py_true, pz_true`	true position relative to LSR	double	Distance (kpc)
`vx_true, vy_true, vz_true`	true velocity relative to LSR	double	Velocity (km s^{-1})
	Photometry		
	true (i.e., after extinction, but before error convolution)		
`phot_g_mean_mag_true`	apparent G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_true`	true apparent G_{BP}-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_true`	true apparent G_{RP}-band mean magnitude	float	Magnitude (mag)
	intrinsic (i.e., before extinction or error convolution)		
`phot_g_mean_mag_int`	apparent G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_int`	intrinsic apparent G_{BP}-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_int`	intrinsic apparent G_{RP}-band mean magnitude	float	Magnitude (mag)
`phot_g_mean_mag_abs`	absolute G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_abs`	absolute G_{BP}-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_abs`	absolute G_{RP}-band mean magnitude	float	Magnitude (mag)
`phot_g_mean_mag_error`	Standard error of G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_error`	Standard error of G_{BP}-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_error`	Standard error of G_{RP}-band mean magnitude	float	Magnitude (mag)
`bp_rp_true`	true $G_{BP} - G_{RP}$ color	float	Magnitude (mag)
`bp_g_true`	true $G_{BP} - G$ color	float	Magnitude (mag)
`g_rp_true`	true $G - G_{RP}$ color	float	Magnitude (mag)
`vmini_true`	true $V - I$ color used for error modeling	float	Magnitude (mag)
	Extinction		
`lognh`	$\log_{10}$ equivalent H column density along line of sight to star	float	surface number density(cm^{-2})
`ebv`	$E(B-V)$ reddening, calculated from N_{H}^{eff}	float	Magnitude (mag)
`A0`	A_0, extinction at 550 nm, assuming $R_V = 3.1$	float	Magnitude (mag)

Table 5.7: (continued)

Quantity	Explanation	Data type	Unit
a_g_val	true line-of-sight extinction in the G band, A_G	float	Magnitude (mag)
e_bp_min_rp_val	true line-of-sight reddening $G_{BP} - G_{RP}$	float	Magnitude (mag)
	Spectroscopy		
radial_velocity_error_corr_factor	correction factor for radial_velocity_error	double	Velocity (km s^{-1})
	Stellar Parameter		
mact	current stellar mass	float	Mass (Solar Mass)
mtip	mass of a star at tip of giant branch for given age, metallicity	float	Mass (Solar Mass)
mini	stellar mass on zero-age main sequence	float	Mass (Solar Mass)
age	$\log_{10}$ of stellar age; identical for all stars generated from the same particle	float	Time (log yr)
teff	stellar effective temperature	float	Temperature (K)
logg	surface gravity	float	Surface Gravity (log cgs)
lum	$\log_{10}$ of stellar luminosity	float	Luminosity (log Solar Luminosity)
	Abundances		
feh	[Fe/H]	float	Magnitude (mag)
alpha	[Mg/Fe]	float	Magnitude (mag)
carbon	[C/H]	float	Magnitude (mag)
helium	[He/H]	float	Magnitude (mag)
nitrogen	[N/H]	float	Magnitude (mag)
sulphur	[S/H]	float	Magnitude (mag)
oxygen	[O/H]	float	Magnitude (mag)
silicon	[Si/H]	float	Magnitude (mag)
calcium	[Ca/H]	float	Magnitude (mag)
magnesium	[Mg/H]	float	Magnitude (mag)

Table 5.7: (continued)

Quantity	Explanation	Data type	Unit
`neon`	[Ne/H]	float	Magnitude (mag)
	Quality Control		
`flag_wd`	flag for potential white dwarfs; see Section 5.2.1	int	

In Table 5.7, we present the column names of the parameters used in the synthetic surveys, as well as their definitions, data types, and units. These column names are categorized by those matching *Gaia* DR2/DR3 (as well as `Ananke` DR2), those that are relating the properties of the simulations (for example the true non-error convolved values), and the properties of the stars (for example their FIRE-2 chemical abundances).

5.4 Use Cases and Limitations

Synthetic surveys can be extremely powerful in testing modeling procedures, calculating false positive rates, and validating methods. This is largely due to the fact that cosmological simulations in general track non-equilibrium dynamics self-consistently, and are therefore powerful tools for exploring dynamical inferences. The need for synthetic surveys is becoming even more critical with the large swaths of data being collected by current and upcoming surveys like *Gaia* and LSST. There are, however, limitations to these studies, based on the nature of the construction of the synthetic surveys. In particular, as was highlighted in Section 5.1, the original simulations have a stellar mass resolution limit of $\sim 7000\,M_{\odot}$. We spawn individual stars from these simulated star particles to generate the mock catalogs. The resulting positions and velocities of the synthetic stars depend on the kernel of choice, limiting the usage of such synthetic surveys. In particular, the internal dynamics of small-scale structures like satellite galaxies and stellar streams are sensitive to the choice of kernel. Studies of small-scale Milky Way structures therefore required careful kernel selection based on the science question at hand; [423], for example, changed the kernel to be able to perform detectability studies of stellar streams with LSST.

More generally, studies of large stellar structures (larger than a few star particles in the original simulation), velocity anisotropies of the Galaxy, and the Milky Way potential, should be robust to the choices of construction of the synthetic surveys, while studies of smaller structures should be treated with care, and potentially a more adequate choice of kernels.

Additionally, the original simulations from which we built these synthetic surveys are not meant to reproduce the Milky Way itself. Indeed, these are cosmological simulations with varying initial conditions, and therefore varying histories; some contain a late merger like m12f, while m12i has a quieter merger history, for example [see e.g. 54, for details of the merger histories of these two galaxies]. Therefore, it is critical to treat these galaxies as examples of galaxies with the same mass as our own, but with their own individual properties. A corollary to this is the dust model adapted for these synthetic surveys is self-consistent with that of the simulations themselves, and therefore is different from that of the Milky Way.

Lastly, synthetic surveys, by definition, are tailor-made to reproduce the observing properties of a particular survey, in the case of *Gaia*, of a particular data release. As new theories, instruments, surveys, and data become available, many key ingredients of synthetic surveys (isochrones, selection functions, error models, and etc.) will become obsolete and inaccurate. To that end, [448] developed `py-ananke`, a universal pipeline designed to generate synthetic surveys from cosmological simulations, adaptable to various instruments. The package builds upon the method employed in S20 and this study, with the crucial feature that allows users to change the ingredients mentioned above. This tool will greatly reduce the time and effort to update synthetic surveys, and ultimately provide a bridge between simulations and observations.

5.5 Conclusion

In this chapter, we presented a new set of synthetic surveys that match *Gaia* DR3, based on the *Latte* suite of the FIRE-2 simulations. This is an update to the synthetic surveys released by [173] that matched the previous data release *Gaia* DR2. These synthetic surveys include three different solar positions for three galaxies. The major changes compared to S20 are an updated set of isochrones matching the latest release, a different treatment of the radial velocity errors that increased the precision of the radial velocity measurements, an update to the proper motion treatment, which decreased the measurement proper motion errors in the synthetic surveys, and an increase in the total number of stars with radial velocity measurements through the update of the selection cuts.

These synthetic surveys are made available to the community on http://ananke.hub.yt/, where they can be used to test any model/analysis pipeline on simulations prior to application to *Gaia* DR3. In particular, these synthetic surveys are the best tool for studies involving the dynamic properties of the Milky Way, especially given that the "true" star particles from the original simulations are also provided. For example, recent studies on the Milky Way rotation curve [e.g., 52] using *Gaia* data have found tentative tension on the dark matter distribution in the Milky Way. Synthetic surveys in this case may be used to quantify the effect of *Gaia* selection function on the accuracy of the measured rotation curve.

More generally, the adoption of synthetic surveys is not only applicable in the study of stars, but also the dynamics of dark matter and properties of gas particles. [160] has made the simulations used in this work publicly available[10], including the formation coordinates of all star particles, as well as catalogs of all satellite galaxies/halos. Therefore, the community can use such information to answer more general questions as to what the field can learn

[10] http://flathub.flatironinstitute.org/fire

through *Gaia* about the galaxy as a whole, from its stellar components as measured through the *Gaia* lens, to the inner workings of the dark matter in the Galaxy that governs the dynamics of the stars.

Chapter 6

Conclusion

In this thesis, I have explored various innovative approaches to uncover the nature of DM and its role in galaxy formation and evolution at small scales, through advanced machine learning techniques and detailed simulations.

Chapter 2 demonstrated how graph neural networks and simulation-based inference can provide stronger constraints on the DM density profiles of dwarf galaxies, addressing the core-cusp discrepancy and other small-scale structure issues in DM halos. In Chapter 3, I introduced FLORAH, a machine-learning framework for generating mass assembly histories of DM halos. This framework successfully recovered key properties such as the time evolution of mass and concentration, and accurately modeled assembly bias. Chapter 4 combined normalizing flows, variational diffusion models, and Transformer-based neural networks to emulate the satellite populations of Milky Way-mass halos under various DM and astrophysics scenarios. In Chapter 5, the *Ananke* synthetic surveys of three Milky Way-like galaxies in the FIRE simulations were presented. These surveys, designed to match observations from the Gaia DR3 catalog, incorporated observational effects and extended previous efforts with Gaia DR2, providing more detailed and accurate synthetic surveys.

The research presented in this thesis highlights the importance of combining advanced computational techniques with detailed simulations to bridge the gap between theoretical models and observational data. Future work will continue to build on these foundations, exploring new DM models, refining machine learning techniques, and further integrating observational data. The ongoing development of synthetic surveys, cosmological simulations, and machine learning techniques will play a crucial role in advancing our knowledge and opening new avenues of research in the study of DM and galaxy formation.

Appendix A

Uncovering dark matter density profiles in dwarf galaxies

A.1 Additional details on the analysis

We elaborate here on several details of the analysis presented in the paper.

A.1.1 Details on the forward model and phase-space distribution function

In this section, we describe in detail the forward model used to generate the dataset. We model the distribution function (DF) using the Osipkov-Merritt (OM) model as proposed by Refs. [305] and [306]. The OM model depends on the angular momentum J and relative energy per unit mass $\mathcal{E} = \phi - v^2/2$, where ϕ is the gravitational potential and v is the velocity, through the variable $Q \equiv \mathcal{E} - J^2/(2r_a)$. Here, r_a is the scale radius of the velocity anisotropy profile $\beta(r) = r^2/(r^2 + r_a^2)$ that defines the transition from an isotropic velocity dispersion at small radii to a radially-biased dispersion at larger radii. To solve for the DF $f(\mathcal{E}, J) = f(Q)$, we first integrate $f(Q)$ over the velocity space to obtain the stellar mass-density profile

$$\rho_\star(r) = \frac{4\pi}{1 + r^2/r_a^2} \int_0^\phi \mathrm{d}Q\, f(Q) \sqrt{2(\phi - Q)}. \tag{A.1}$$

Note that $f(Q) = 0$ for $Q < 0$. The final DF $f(Q)$ can be obtained by Abel transforming $\rho_\star$,

$$f(Q) = \frac{1}{2\pi\sqrt{2}} \frac{\mathrm{d}G(Q)}{\mathrm{d}Q}, \quad \text{where} \quad G(Q) = -\int_0^Q \frac{\mathrm{d}\rho_Q}{\mathrm{d}\phi} \frac{\mathrm{d}\phi}{\sqrt{Q - \phi}} \quad \text{and} \quad \rho(Q) = \left(1 + \frac{r^2}{r_a^2}\right) \rho_\star(r). \tag{A.2}$$

Assuming that the system is dominated by DM, the gravitational potential ϕ depends on the DM density profile via Poisson's equation $\nabla^2\phi = -4\pi G\rho_{\rm DM}(r)$. The DM density profile is parameterized through the generalized Navarro-Frenk-White model as in Chapter 2:

$$\rho_{\rm DM}^{\rm gNFW}(r) = \rho_0 \left(\frac{r}{r_s}\right)^{-\gamma} \left(1+\frac{r}{r_s}\right)^{-(3-\gamma)}, \tag{A.3}$$

where ρ_0 is the density normalization, r_s is the DM scale radius, and γ is the inner density slope. The stellar density profile is

$$\rho_\star(r) = \rho_{0,\star}\left[1+\frac{r^2}{r_\star^2}\right]^{-5/2}, \tag{A.4}$$

where $r_\star$ is the stellar scale radius and $\rho_{0,\star}$ is the density normalization of the stellar profile. For $\rho_{0,\star} = 3M/4\pi r_\star^3$ and $M \propto L$, where M and L is the enclosed mass and luminosity, this is proportional to the 3-D Plummer profile

$$\nu(r) = \frac{3L}{4\pi r_\star^3}\left(1+\frac{r^2}{r_\star^2}\right)^{-5/2} \tag{A.5}$$

which we use to characterize our stellar profile in Chapter 2. The DF $f(Q)$ may be written in terms of the radius r and the tangential and radial velocity v_r and v_t (i.e. $f(Q) = f(r, v_r, v_t)$) since $J = rv_t$ and $\mathcal{E} = \phi(r) - (v_t^2 + v_r^2)/2$. `StarSampler` samples the coordinates r, v_r, v_t of each star from $f(r, v_r, v_t)$ using importance sampling [301]–[303]. The 6-D coordinates (x, y, z, v_x, v_y, v_z) are then calculated by assuming a random projection direction and spherical symmetry.

A.1.2 Details on Jeans analysis procedure

In this section, we briefly outline the Jeans analysis procedure used in this work, closely following Ref. [254]. Following the derivation from Refs. [449] and [450], we first assume the system follows the collisionless Boltzmann equation

$$\frac{\partial f}{\partial t} + \vec{v}\frac{\partial f}{\partial \vec{x}} - \frac{\partial \Phi}{\partial \vec{x}}\cdot\frac{\partial f}{\partial \vec{v}} = 0, \tag{A.6}$$

where ϕ is the gravitational potential, $f = f(\vec{x}, \vec{v})$ is the phase-space distribution function, and $(\vec{x}, \vec{v})$ are the phase space coordinates of tracer stars.

Working in the dwarf galaxy's spherical coordinate system (r, θ, ϕ) and assuming spherical symmetry and steady state, we multiply Eq. (A.6) by the radial velocity v_r and integrate

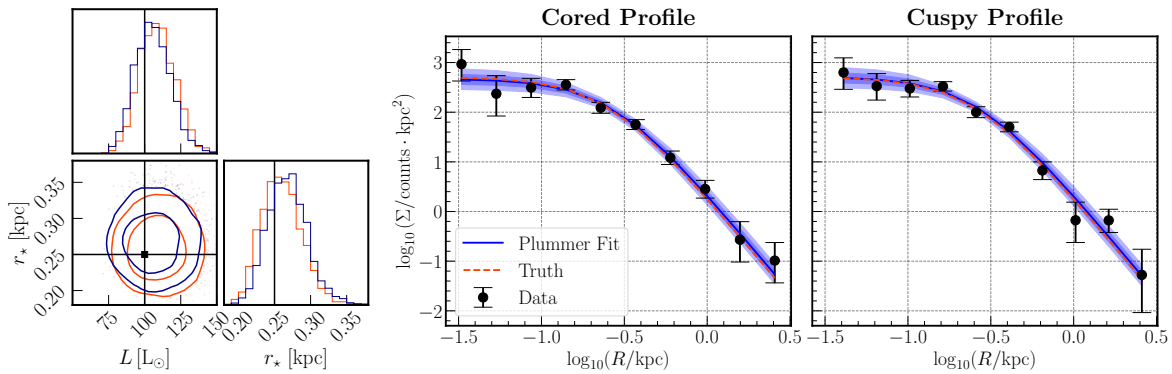

Figure A.1: The initial Plummer fit for the two test galaxies presented in the result of Chapter 2. The left panel shows the 68% and 95% credible intervals of the posterior distributions of L and $r_\star$. Both galaxies have the same light profile ($L = 100\,\mathrm{L}_\odot$ and $r_\star = 0.25$ kpc) but different DM profiles: cored DM profile (red) versus cuspy DM profile (blue). The right panel shows the best fit, the truth Plummer profile, and the binned data with Poisson uncertainties for both galaxies. The dark blue and light blue bands show the middle-68% and 95% credible intervals of the reconstructed profile.

over all velocity components to obtain the spherical Jeans equation

$$\frac{1}{\nu}\left[\frac{\partial}{\partial r}(\nu\sigma_r^2) + \frac{2\beta(r)}{r}(\nu\sigma_r^2)\right] = -\frac{\partial\phi}{\partial r} = -\frac{GM(<r)}{r^2}, \tag{A.7}$$

where $\nu = \int \mathrm{d}^3\vec{v}\, f(\vec{x},\vec{v})$ is the number density of the tracer stars, σ_i is the velocity dispersion $\sigma_i = \sqrt{\langle v_i^2\rangle - \langle v_i\rangle}$ for $i \in (r,\theta,\phi)$, and $\beta(r) = 1 - (\sigma_\theta^2 + \sigma_\phi^2)/(2\sigma_r^2)$ is the velocity anisotropy profile. The gravitational potential ϕ is assumed to be dominated by DM and may be written as $\phi = -GM(<r)/r$, where G is the gravitational constant and $M(<r)$ is the enclosed mass of DM. The Jeans equation (A.7) has therefore the solution

$$\sigma^2(r)\nu(r) = \frac{1}{g(r)}\int_r^\infty \frac{GM(<r')\nu(r')}{r'^2} g(r')\mathrm{d}r' \quad \text{where} \quad g(r) = \exp\left(2\int_0^r \frac{\beta(r')}{r'}\mathrm{d}r'\right). \tag{A.8}$$

Projecting Eq. (A.8) along the line of sight using the Abel transformation $s(r) \to S(R)$ for the spherically symmetric function $s(r)$,

$$S(R) = 2\int_R^\infty \frac{s(r)r\,\mathrm{d}r}{\sqrt{r^2 - R^2}}, \tag{A.9}$$

we obtain

$$\sigma_p^2(R)I(R) = 2\int_R^\infty \left(1 - \beta(r)\frac{R^2}{r^2}\right)\frac{\nu(r)\sigma_r^2(r)r}{\sqrt{r^2 - R^2}}\mathrm{d}r, \tag{A.10}$$

where σ_p is the projected velocity dispersion profile and $I(R)$ is the projected number density

Table A.1: Prior ranges for the DM and stellar parameter for the Jeans analysis and our method.

	Parameter	**Prior distribution**
DM density profile	$\log_{10}(\rho_0/(\mathrm{M}_\odot\,\mathrm{kpc}^{-3})$	$\mathcal{U}(5,8)$
	$\log_{10}(r_s/\mathrm{kpc}))$	$\mathcal{U}(-1,0.7)$
	γ	$\mathcal{U}(-1,2)$
Light profile (GNN + Flows)	$r_\star/r_s$	$\mathcal{U}(0.2,1)$
	$r_a/r_\star$	$\mathcal{U}(0.5,2)$
Light profile (Jeans analysis)	$\log_{10}(L/\mathrm{L}_\odot)$	$\mathcal{U}(-2,5)$
	$\log_{10}(r_\star/\mathrm{kpc})$	$\mathcal{U}(-3,3)$
	$r_a/r_\star$	$\mathcal{U}(0.5,2)$
	$\bar{v}/(\mathrm{km\,s}^{-1})$	$\mathcal{U}(-100,100)$

of tracer stars, also known as the surface brightness or the light profile. In our analysis, we parameterize the DM density profile using the gNFW profile in Eq. (A.3). The light profile $I(R)$ is the projection along the line-of-sight of the 3-D Plummer profile in Eq. (A.5) and is given by

$$I(R) = \frac{L}{\pi r_\star^2}\left(1+\frac{R^2}{r_\star^2}\right)^{-2}, \tag{A.11}$$

which has two free parameters: the total luminosity L and the scale length $r_\star$.

The Jeans analysis requires two separate fits: fitting the light profile and subsequently fitting the DM profile. We first fit L and $r_\star$ using the priors in Table A.1. We approximate the posteriors as Gaussian distributions and use them as the priors for L and $r_\star$ in the second fit. To compare the performance between Jeans analysis and our machine learning framework, the DM priors are set to be the same as those used to generate the training datasets for the GNN and normalizing flow method, and are summarized in Table A.1.

For the initial fit of the light profile, we bin the data in $\log_{10}$-spaced bins in the projected radius R. The number of bins is chosen to be $\sim\sqrt{N_{\rm stars}}$ where $N_{\rm stars}$ is the number of stars. Similarly to Ref. [254], we assume Poisson uncertainties on the number of stars in each bin. Let $n_i(\theta)$ and $\hat{n}_i$ be the predicted and mean number of stars (with θ is the parameters of the

light profile model) in i^{th} bins. We construct the binned likelihood [451]

$$\ln \mathcal{L}_{\text{Plummer}} = -\frac{1}{2} \sum_{i=1}^{N_{\text{stars}}} \frac{(\hat{n}_i - n_i(\theta))^2}{V_i - V_i'(\hat{n}_i - n_i(\theta))}, \tag{A.12}$$

where $V = \sigma_{\text{lo}}\sigma_{\text{hi}}$ and $V' = \sigma_{\text{hi}} - \sigma_{\text{lo}}$; here, σ_{lo} and σ_{hi} are the asymmetric Poisson uncertainties. We refer to Ref. [254] for further details. As described in Chapter 2, we sample the posterior with nested sampling [316], [317] using the DYNESTY module [318]. We use $n_{\text{live}} = 500$ live points and a convergence tolerance of $\Delta \ln \mathcal{Z} = 0.1$ on the estimated remaining contribution to the log-evidence. In Fig. A.1, we show the initial Plummer fit for the two test galaxies presented in the result of Chapter 2. The posteriors for L and $r_\star$ are well-constrained and agree well with the true Plummer profile.

As mentioned, we approximate the posteriors of L and $r_\star$ as Gaussian distributions and use them as priors in the second fit of the DM profile. Unlike the initial Plummer fit, for the Jeans analysis, we construct an unbinned Gaussian likelihood. The likelihood is given by Ref. [189],

$$\mathcal{L}_{\text{Jeans}} = \prod_{i=1}^{N_{\text{stars}}} \frac{(2\pi)^{-1/2}}{\sqrt{\sigma_p^2(R_i) + \Delta {v_i}^2}} \exp\left[-\frac{1}{2}\left(\frac{(v_i - \bar{v})^2}{\sigma_p^2(R_i) + \Delta {v_i}^2}\right)\right], \tag{A.13}$$

where $\sigma_p(R)$ is the projected velocity dispersion profile in Eq. (A.10), $\bar{v}$ is the mean velocity of tracer stars, and v_i and Δv_i is the line-of-sight velocity and its measurement error for star i. We treat the mean velocity $\bar{v}$ as a nuisance parameters and fit it together with the DM and light profile, using the prior distribution in Table A.1. We sample the joint DM and light profile posteriors using DYNESTY [318] with the same sampler configuration as in the initial Plummer fit.

A.1.3 Details on the *J*-factor calculation

In this section, we outline the details on the annihilation J-factor calculation. The J-factor is defined as the integral along the line-of-sight s and over solid angle Ω of the squared DM density corresponding to a source target,

$$J = \int \mathrm{d}s \int \mathrm{d}\Omega \, \rho^2(s, \Omega). \tag{A.14}$$

We transform the coordinate system in the Eq. (A.14) from a spherical coordinate system centered at the Earth's location to a spherical coordinate system centered at the dwarf

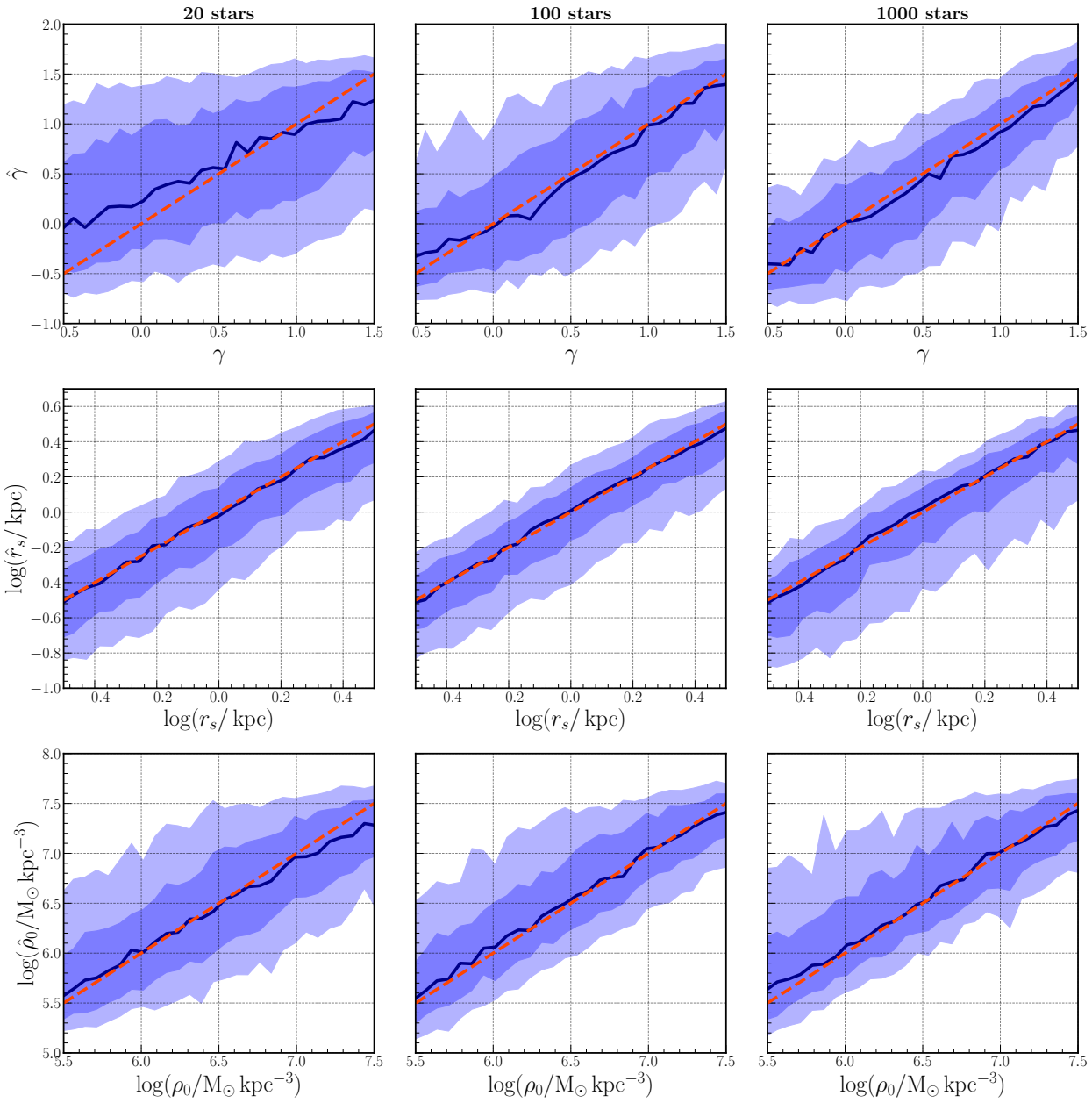

Figure A.2: The predicted DM parameters versus truth DM parameters for three different runs. In each run, a GNN and normalizing is trained and tested on galaxies with a mean number of stars $\mu_{\mathrm{stars}} = 20\,(\mathrm{left}), 100\,(\mathrm{center}), 1000\,(\mathrm{right})$ stars. The line-of-sight velocity measurement error is set to be $\Delta v = 0.1\,\mathrm{km/s}$. The $\mu_{\mathrm{stars}} = 100$ case (middle column) is the same case as shown in the result of Chapter 2. The median (solid blue line), middle-68% percentile (blue band), and middle-95% (light blue band) containment regions of each bin are shown. The diagonal dashed red line shows the locus of equality between true and predicted values.

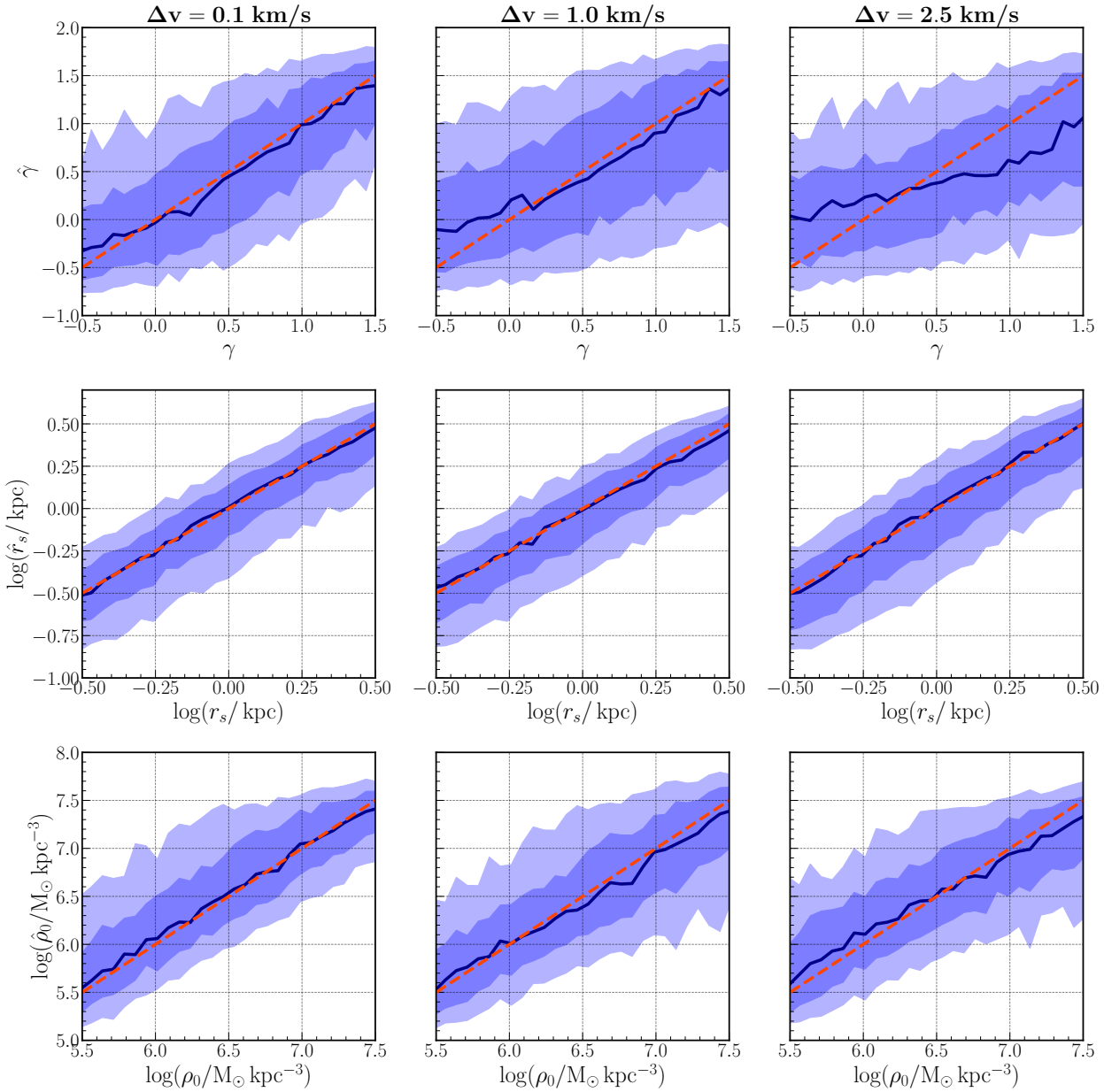

Figure A.3: Same as Fig. A.2, but with different line-of-sight velocity errors Δv. The mean number of stars μ_{stars} is 100 stars. The $\Delta v = 0.1\,\mathrm{km/s}$ case (left column) is the same case as shown in the result of Chapter 2.

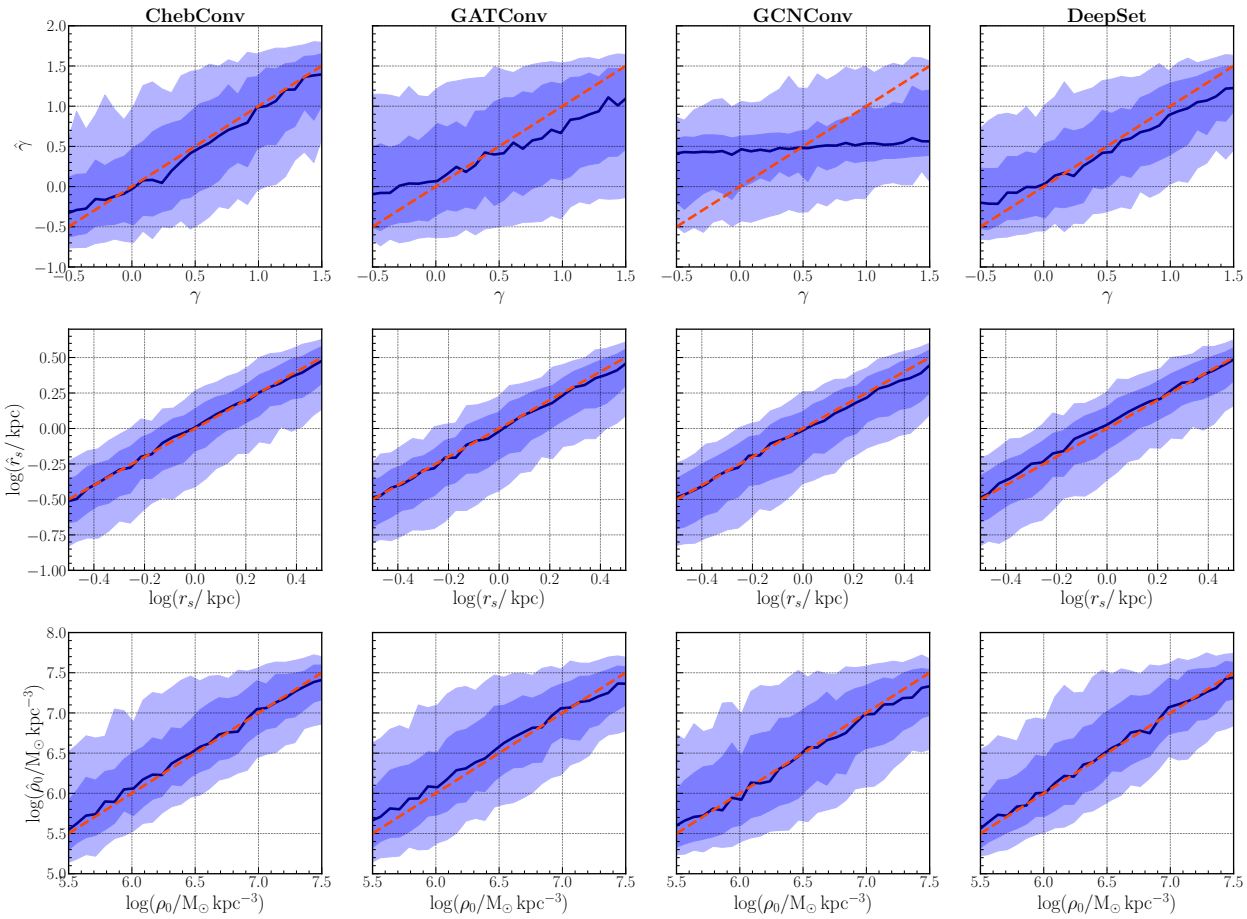

Figure A.4: Same as Fig. A.2, but with variation in the type of graph convolutional layer. Each network is trained and tested on the dataset presented in Chapter 2, i.e. $\mu_{\mathrm{stars}} = 100$ stars and $\Delta v = 0.1\,\mathrm{km/s}$. The ChebConv case (left column) refers to the baseline graph-convolutional scheme presented in Chapter 2.

galaxy considered. Let $\vec{r'}$ be the position from the center of the galaxy; the J-factor in this coordinate system is given by

$$J = \int \mathrm{d}s \int \mathrm{d}\Omega\, \rho^2(s,\Omega) = \int \mathrm{d}V \frac{\rho^2(s,\Omega)}{s^2} = \int \mathrm{d}V' \frac{\rho^2(r',\Omega')}{r'^2 - 2d_c r' \cos\theta' + d_c^2}, \tag{A.15}$$

where $d_c = |\vec{r}|$ is the comoving distance and $\vec{r}\cdot\vec{r'} = d_c r' \cos\theta'$. Note that the volume elements between the two coordinate systems are equal, $\mathrm{d}V = \mathrm{d}V'$, since the transformation is only a translation. We integrate Eq. (A.15) up to the virial radius $r_{\rm vir}$, defined as the radius within which the mean density of the halo is equal to a specified overdensity factor times the critical density of the Universe, $\overline{\rho}(r < r_{\rm vir}) = \Delta_{\rm vir}\rho_c$. As per common convention, we take $\Delta_{\rm vir} = 200$. We assume the distant-source approximation $d_c \gg r_{\rm vir} > r'$, justified for most Milky Way dwarf galaxies, which allows us to write J as a volume integral in spherical coordinates [452]

$$J = \int \mathrm{d}V' \frac{\rho^2(r',\Omega')}{r'^2 - 2d_c r' \cos\theta' + d_c^2} \approx \frac{1}{d_c^2} \int \mathrm{d}V'\, \rho^2(r'). \tag{A.16}$$

Plugging the gNFW density profile Eq. (A.3) into the above expression, we find

$$J = \frac{1}{d_c^2} \int \mathrm{d}V' \, (\rho_{\rm DM}^{\rm gNFW})^2(r') = \frac{4\pi\rho_s^2 r_s^2}{d_c^2} \int_0^{r_{\rm vir}} \mathrm{d}r' \left(\frac{r}{r_s}\right)^{2-2\gamma} \left(1 + \frac{r}{r_s}\right)^{2\gamma-6} \tag{A.17}$$

$$= \frac{4\pi\rho_s^2 r_s^3}{d_c^2} \left[\frac{c_{\rm vir}^{3-2\gamma}(1+c_{\rm vir})^{2\gamma-5}\,(20 + 4\gamma^2 + 2c_{\rm vir}(5+c_{\rm vir}) - 2\gamma(9+2c_{\rm vir}))}{(5-2\gamma)(4-2\gamma)(3-2\gamma)} \right] \tag{A.18}$$

where $c_{\rm vir} \equiv r_{\rm vir}/r_s$ is the virial concentration of the dwarf galaxy. Note that the integration only converges when $\gamma < 1.5$, which is the case for the DM profiles shown in Chapter 2.

A.1.4 Prior distributions on the parameters of interest

The prior distributions for both our method and the Jeans analysis are shown in Table A.1. The priors on the DM density profile parameters are the same between the Jeans analysis and our method. As described in Section A.1.2, the Jeans analysis consists of an initial fit of the light profile parameters L and $r_\star$, and a subsequent joint fit of both the DM and light profile parameters.

A.2 Additional results

A.2.1 Systematic variations on the analysis

We explore variations on the assumptions made in our baseline analysis, including the effect of varying the mean number of stars $\mu_{\rm stars}$ and line-of-sight velocity measurement error Δv. We also examine the performance of different graph convolution schemes.

Variations on the number of stars: We generate three datasets (including the one presented in Chapter 2) with the same velocity measurement error $\Delta = 0.1\,\mathrm{km/s}$ and different mean number of stars: $\mu_{\rm stars} = 20, 100, 1000$. As in the baseline case, each dataset contain 80,000 training samples, 10,000 validation samples, and 10,000 test samples. For each dataset, we train a GNN and normalizing flow with the same hyperparameters as described in Chapter 2 and plot the results on the test samples in Fig. A.2. We take the marginal medians as the predicted DM parameters, bin them based on their truth values, and show the median (solid blue line), middle-68% (blue bands), and middle-95% (light blue bands) containment regions. The baseline $\mu_{\rm stars} = 100$ stars (middle column) case is the same as that in Chapter 2 and is shown here for comparison.

The performance on the scale radius r_s (middle row) and density normalization ρ_s (bottom row) is similar for all three cases. For the inner slope γ (top row), we see that increasing the number of stars helps increase the accuracy of the marginal medians. This is expected because the underlying phase-space distribution is more completely sampled for a larger number of observed tracer stars. We observed a similar increase in prediction accuracy with increasing sample size in Ref. [254]. The marginal medians for the $\mu_{\rm stars} = 20$ case are slightly biased towards the tails of the γ prior distribution. This is potentially due to clipping caused by a finite prior range that prevents the marginal median from being centered around the true value when constraining power is low.

Variations on measurement uncertainty: We generate three datasets with the same mean number of stars $\mu_{\rm stars}$ and different velocity measurement errors $\Delta v = 0.1,\ 1.0, 2.5\,\mathrm{km/s}$. Similar to our baseline case of $\Delta v = 0.1\,\mathrm{km/s}$, each dataset has 80,000 training samples, 10,000 validation samples, and 10,000 test samples, and we train our pipeline on each dataset separately. The results of this variation are shown in Fig. A.3 in the same format as Fig. A.2 (with the baseline $\Delta v = 0.1\,\mathrm{km/s}$ case the same as in the result of Chapter 2).

Again, we observe that for the scale radius r_s (middle row) and the density normalization ρ_0 (bottom row), the performance is approximately constant across all variations. Similar to

Ref. [254], as the measurement uncertainties increase we observe a decrease in constraining power of the central slope γ. This is to be expected because we did not explicitly account for the uncertainties in the neural network architecture—sampling from the noise model is treated as a form of data augmentation, with larger error magnitudes leading to increased sample variance and reduced sample efficiency. We defer the explicit inclusion of observational uncertainties in the neural network construction to future work.

Variations on the graph convolution scheme: In Fig. A.4 we show variations on the type of graph convolutional layer used, otherwise keeping all hyperparameters (e.g., channel dimension) the same as the baseline case. This case, labeled 'ChebConv' and shown in the left-most column, uses the graph convolution prescription from Ref. [308].

The second column shows results using the graph attention layer from Ref. [242], which uses the attention mechanism to implicitly weigh neighboring nodes. The final column shows results using the deep set architecture, where node-wise features are obtained using a dense network and aggregated through averaging, before finally being passed through a dense layer as in the baseline case. Relatively good recovery of all parameters can be seen in these cases. Together, these results point to the fact that aggregation of neighborhood information may not play a key role in the success of our method on the examples tested. We expect this fact to not hold on more realistic test cases—in particular, the systems in this paper were chosen to be relatively simple (spherical and dynamically equilibrated) in order to enable a direct comparison with the conventional Jeans analysis method. With these assumptions and our choice of stellar and DM profiles, the node-wise features $\{R, v_{\rm los}\}$ can be assumed to be independent.

Finally, the third column of Fig. A.4 shows results using the graph convolutional layer from Ref. [453] with the default configuration in PyTorch Geometric, showing poor recovery of the inner slope γ.

A.2.2 Comparison of inner density slopes

In order to quantitatively compare the inferred inner density slope γ posteriors for a large galaxy sample, we compute and summarize the Jensen-Shannon (JS)-divergence [454], [455] between 100 pairs of cored and cuspy γ posteriors using our method as well as Jeans analysis. Given two probability distribution functions $P(x)$ and $Q(x)$, defined over the same probability space $\mathcal{X} \ni x$, the Jensen-Shannon (JS)-divergence is defined as

$$D_{\rm JS}(P||Q) \equiv \frac{1}{2}\left[D_{\rm KL}(P||M) + D_{\rm KL}(Q||M)\right], \tag{A.19}$$

where $M(x) = P(x) + Q(x)$ and $D_{\rm KL}$ is the Kullback–Leibler (KL) divergence, for which we use the definition

$$D_{\rm KL}(P||Q) = \sum_{x\in\mathcal{X}} P(x)\log_2\left(\frac{P(x)}{Q(x)}\right). \tag{A.20}$$

The KL divergence of P from Q represents the expected entropy gain from using Q as an approximation for truth distribution P [456]. The JS-divergence is based on the KL-divergence but has more desirable properties for the present case, specifically being symmetric in the two posterior distributions. Note that using a $\log_2$ definition of the constituent KL-divergences, the JS-divergence is constrained to lie within the range $D_{\rm JS} \in [0, 1]$, with 0 corresponding to two identical distributions and 1 two non-overlapping distributions.

We generate 100 galaxies with cored DM profiles ($\gamma = 0$) and 100 galaxies with cuspy profiles ($\gamma = 1$). These galaxies have the same DM scale radius $r_s = 1\,\mathrm{kpc}$, density normalization $\rho_0 = 10^7\,\mathrm{M}_\odot/\mathrm{kpc}^3$, and light profile parameters as the two galaxies presented in the result of Chapter 2. We obtain samples the DM parameter posteriors for each galaxy and calculate $D_{\rm JS}$ using Eq. (A.19) for each pair of cored and cuspy galaxies (in total, there are 10,000 pairs). We show the distribution of the JS-divergences $D_{\rm JS}$ between the $\gamma = 0$ posterior and $\gamma = 1$ posterior for the Jeans analysis and for our method in Fig. A.5. The median and middle-68% containment region values are $D_{\rm JS}^{\rm Jeans} = 0.417^{+0.288}_{-0.280}$ and $D_{\rm JS}^{\rm GNN} = 0.629^{+0.196}_{-0.278}$. Our method generically produces higher values of $D_{\rm JS}$ compared to the Jeans analysis, corresponding to a larger contrast between the cored and cuspy γ posteriors and thus increased ability to distinguish between the two scenarios given a set of observations.

A.2.3 Comparison of the inferred *J*-factors

As mentioned in Chapter 2, we calculate the J-factors (normalized to distance of 100 kpc) for 100 galaxies randomly sampled from our test set. In this section, we plot the posterior distribution of the J-factor for a few examples for the Jeans analysis and our method in Fig. A.6. In all cases shown, the GNN and normalizing flow can predict the J-factor with a similar or higher accuracy compared to the Jeans analysis.

A.2.4 Test of statistical coverage of the inferred posteriors

Simulation-based inference methods such as those employed in this work can be susceptible to producing overconfident posteriors [457]. In this section, we examine the quality of the posterior distributions produced by our simulation-based inference pipeline following the prescription in Ref. [457].

Using the same notation as in Chapter 2, let θ be the parameters of interest (i.e. the

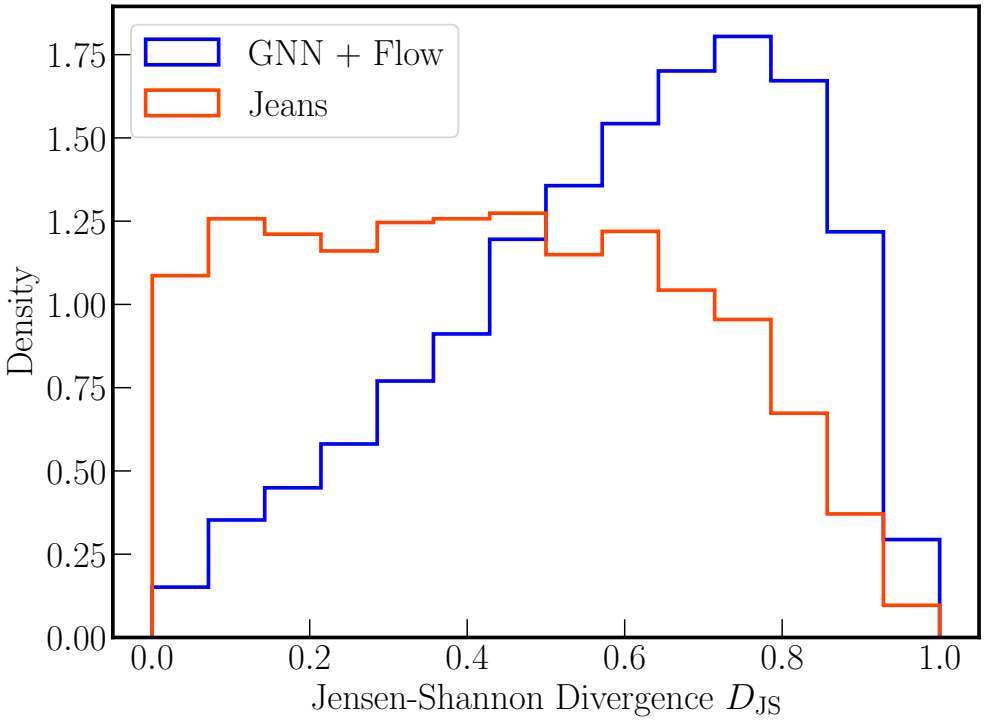

Figure A.5: The JS divergence $D_{\rm JS}$ between the $\gamma = 0$ and $\gamma = 1$ posteriors for the Jeans analysis (red) and the GNN and normalizing flow model (blue). As evidenced from the higher values of the $D_{\rm JS}$, compared to the Jeans analysis, the GNN and normalizing flow produces $\gamma = 0$ and $\gamma = 1$ more distinct from each other.

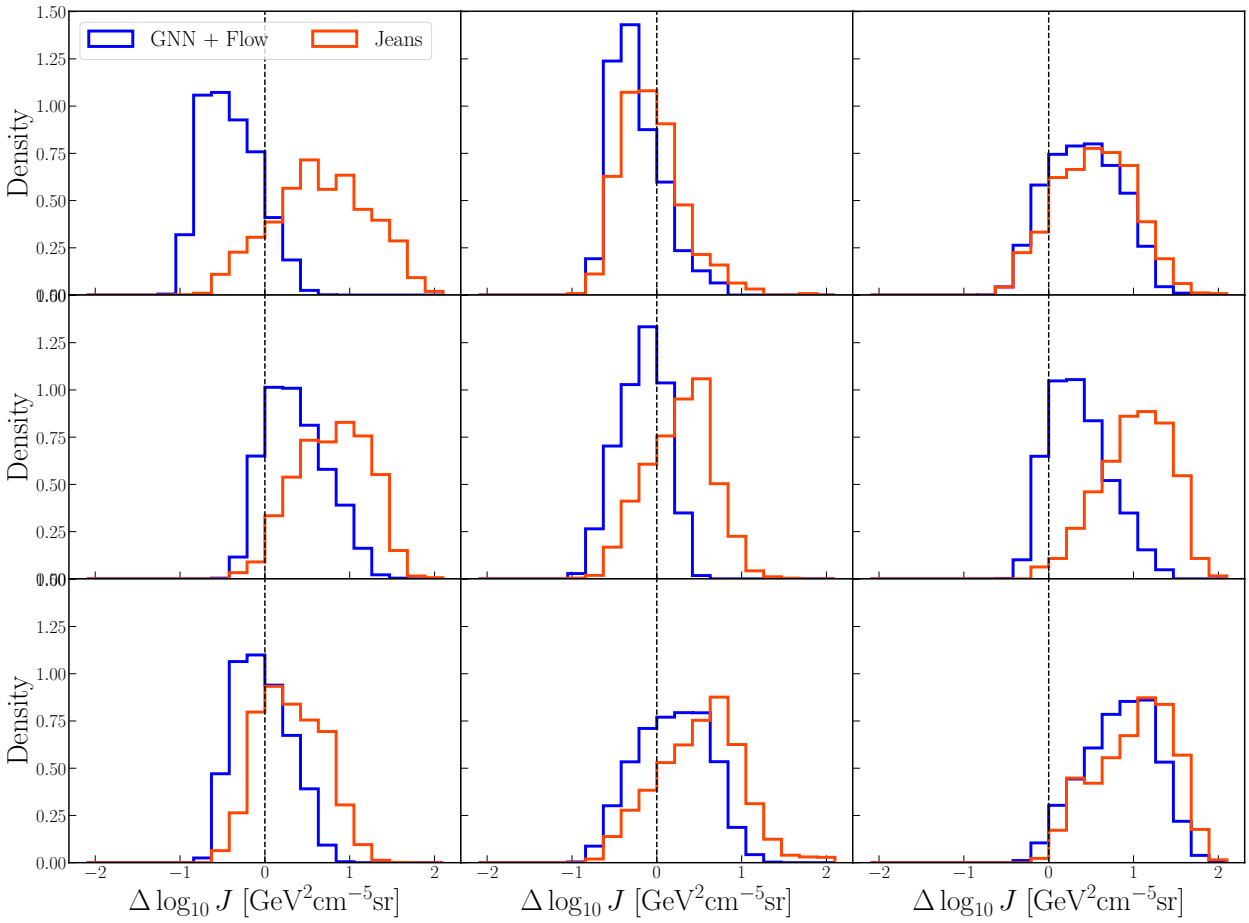

Figure A.6: Posterior distributions of the J-factors predicted by the Jeans analysis (red) and by the GNN and normalizing flow (blue). Each panel show the $\Delta \log_{10} J = \log_{10} J_{\text{predict}} - \log_{10} J_{\text{truth}}$ posteriors of an example galaxy. The dashed black line represents the truth J-factor value (i.e. $\Delta \log_{10} J = 0$).

DM and stellar parameters) and x be the observable (i.e. the graph representation of a galaxy constructed from its the stellar kinematics). We denote our learned posterior density estimator as $\hat{p}(\theta|x)$. For a confidence level $1-\alpha$, the expected coverage probability is

$$\mathbb{E}_{(\theta,x)\sim p(\theta,x)} \left[\mathbb{1}_\Theta(\theta \in \Theta_{\hat{p}(\theta|x)}(1-\alpha)) \right], \tag{A.21}$$

where $\Theta_{\hat{p}(\theta|x)}(1-\alpha)$ gives the $1-\alpha$ highest posterior density interval (HDPI) of the estimator $\hat{p}(\theta|x)$ and $\mathbb{1}_\Theta()$ is an indicator function mapping samples that fall within the HDPI to unity. Given N samples from the joint distribution $(\theta^\star, x) \sim p(\theta, x)$, the empirical expected coverage

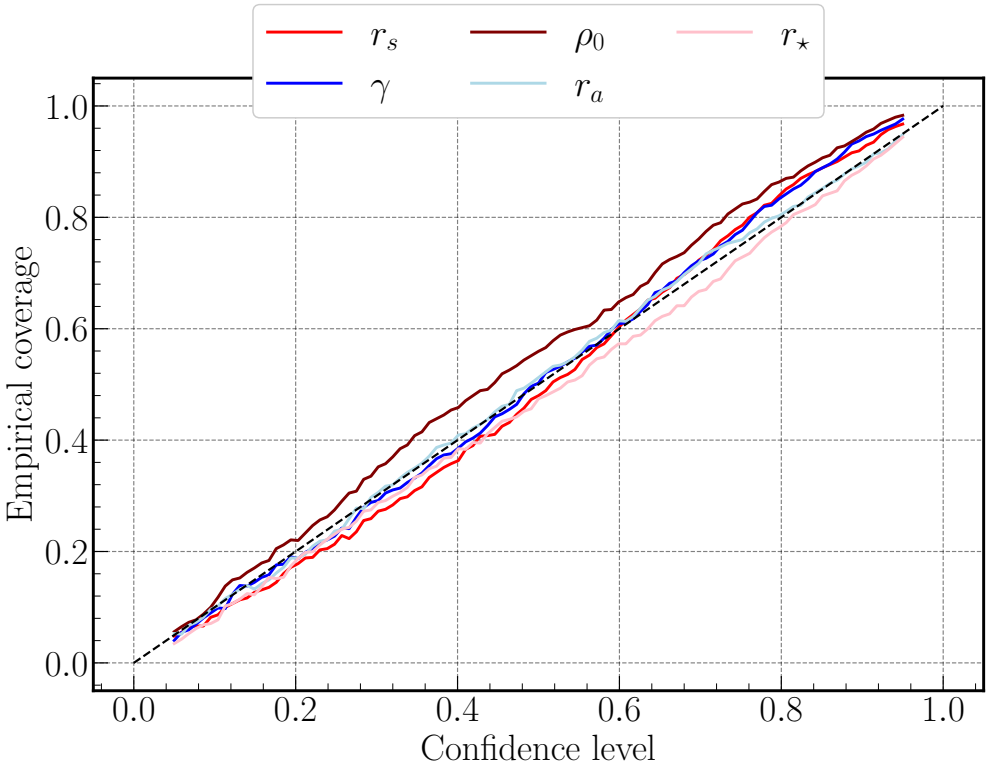

Figure A.7: The expected coverage of the marginal DM and stellar parameters posteriors by the model presented in Chapter 2 ($\mu_{\rm stars} = 100, \Delta v = 0.1\,{\rm km/s}$). If an estimator produces perfectly calibrated posteriors, then its empirical expected coverage probability is equal to the nominal expected coverage probability (dashed diagonal black line). The estimator is conservative (overconfident) if it produces an empirical expected coverage probability above (below) the diagonal line.

for the posterior estimator $\hat{p}(\theta|x)$ is defined as

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_\Theta\left(\theta^\star \in \Theta_{\hat{p}(\theta|x)}(1-\alpha)\right). \tag{A.22}$$

The nominal expected coverage is the expected coverage in the case when $\hat{p}(\theta|x) = p(\theta|x)$ and equals to the confidence level $1-\alpha$. In general, we want our estimator to have an empirical expected coverage probability larger than or equal to the nominal expected coverage probability at all confidence levels. Such estimators will produce conservative posteriors, in contrast to overconfident posteriors which may spuriously exclude allowable regions of parameter space.

In Fig. A.7, we plot the empirical expected coverage probability for the marginal posteriors as produced by the baseline model presented in Chapter 2 (i.e. $\mu_{\rm stars} = 100$ and $\Delta v = 0.1\,{\rm km/s}$) against the nominal confidence levels. The dashed diagonal black line repre-

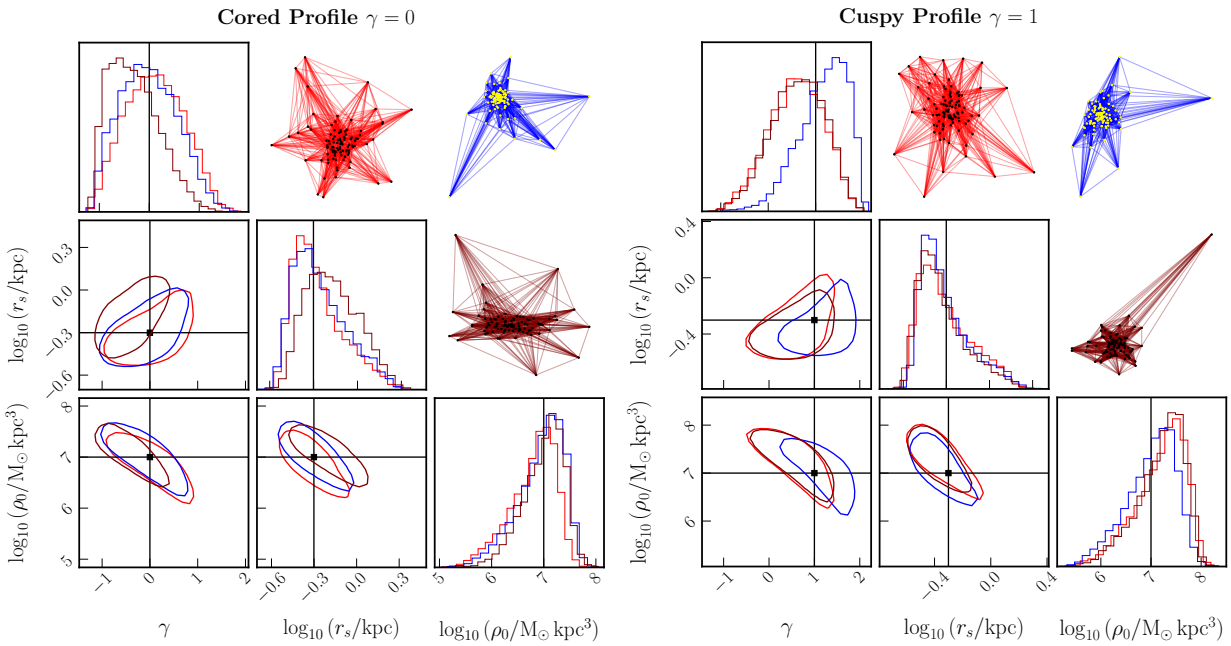

Figure A.8: Example corner plots of the posterior DM parameters of two test galaxies, each with three orthogonal projections. The left (right) panel shows the posteriors for a galaxy with a cored (cuspy) DM profile. The contour lines show the 68-% containment region, with each color representing a different projection. For each projection, its graph representation is also shown (with the edge color matching the contour color) on the upper right corner.

sents the nominal expected coverage probability. A conservative estimator will lie above the diagonal line, while an overconfident estimator will lie below it. In general, the posteriors produced by our model lie very close to the well-calibrated regime, although the posteriors for the stellar radius $r_\star$ are slightly overconfident. We note that methodological improvement in calibration quality of posteriors produced using forward-modeling approaches is an ongoing, active area of research [458].

A.2.5 Robustness to observational projection

Typically, only line-of-sight velocities and angular coordinates of tracer stars are observationally accessible, presenting the challenge of working with incomplete phase-space information when inferring the DM density profile. A test of our method is then its susceptibility to the specific direction from which a dwarf galaxy is viewed—its observational projection. Since different projections correspond to the same latent DM parameters, it is desirable for different projections to produce similar summary representations, and therefore similar posterior distributions. Since this is not explicitly baked into the network, approximate projective symmetry can be learned implicitly using the training sample.

We take the galaxies with the same DM profiles presented in Chapter 2 and project them onto orthogonal planes. For each galaxy, we thus obtain 3 orthogonal projections (including the original projection). In Fig. A.8, we show the 68% contour line of the posterior DM parameters for three orthogonal projections of two galaxies (one with a cored DM profile and one with a cuspy DM profile) and the graph representations of the projections. For each projection, its graph representation matches the color of the contour line of the DM posteriors. It can be seen that even though the graph representations may vary significantly between projections (e.g. the positions of each node may shift, forming new edge connections or breaking up old ones), the DM posteriors remain consistent. Note that we do not expect the DM posteriors to be identical, since the information given to the GNN is not the identical between projections.

Appendix B

Emulating subhalo populations under alternative dark matter scenarios with Diffusion Models

B.1 Normalizing flows

Normalizing flows model the distribution $p(\mathbf{y})$ of the data $\mathbf{y}$ by transforming it from a base distribution $p(\mathbf{u})$ via an invertible transformation. Let f be an invertible transformation that maps a random variable $\mathbf{u}$ to $\mathbf{y}$, i.e. $\mathbf{y} = f(\mathbf{u})$. Using the change of variables formula, we can write the $p(\mathbf{y})$ as:

$$p(\mathbf{y}) = p(f^{-1}(\mathbf{y})) \left| \det \left(\frac{\partial f^{-1}}{\partial \mathbf{y}} \right) \right| \tag{B.1}$$

where the last term is the absolute value of the determinant of the Jacobian matrix of f^{-1} at $\mathbf{y}$ and accounts for the change in volume under the transformation. The base distribution $p(\mathbf{u})$ is often set to a simple, well-understood distribution such as a standard Gaussian. The transformation f is a neural network with trainable parameters but with a few restrictions. First, we emphasize the invertibility of the transformation, which allows for both sampling (forward transformation from $p(\mathbf{u})$ to $p(\mathbf{y})$) and density estimation (inverse transformation from $p(\mathbf{y})$ to $p(\mathbf{u})$). In addition, the Jacobian determinant must be tractable to ensure the model can be efficiently trained and applied to high-dimensional data. These requirements restrict the type of transformations available, with common flows being the masked autoregressive flows [370], [459] and neural spline flows [387].

In practice, the normalizing flows are constructed from a sequence of T discrete transformations $f = f_1 \circ f_2 \circ ... \circ f_T$, where each transformation f_i is also invertible and has tractable

Jacobian determinant. The data likelihood $p(\mathbf{y})$ can be written as:

$$p_T(\mathbf{y}) = p_0(\mathbf{u}_0) \prod_{i=1}^{T} \left| \det \left(\frac{\partial f_i^{-1}}{\partial \mathbf{u}_{i-1}} \right) \right| \tag{B.2}$$

where $\mathbf{y} = \mathbf{u}_T$. The optimization objective is thus simply the negative log-likelihood $\log p_T(\mathbf{y})$, i.e.

$$\mathcal{L}_{\text{flows}}(\mathbf{y}) = -\log p_0(f^{-1}(\mathbf{y})) - \sum_{i=1}^{T} \log \left| \det \left(\frac{\partial f_i^{-1}}{\partial \mathbf{u}_{i-1}} \right) \right| . \tag{B.3}$$

Each transformation f_i is parameterized by a neural network with parameters ϕ_i. During training, the full set of parameters $\phi = \{\phi_1, \phi_2, ..., \phi_T\}$ is optimized simultaneously. In the case of a conditional generative model, with conditioning features $\boldsymbol{\theta}$, we simply include $\boldsymbol{\theta}$ into each transformation, i.e. $f_i(\mathbf{y}) \rightarrow f_i(\mathbf{y}, \boldsymbol{\theta})$.

B.2 Diffusion Model

In this Appendix, we present a summary of the inner workings of variational diffusion models (VDM), following the discussion in [235]. For more detailed mathematical formalism and derivation, we also referred readers to a review in [234].

B.2.1 Forward diffusion

In the context of a Gaussian diffusion model, forward diffusion refers to the process of gradually corrupting an original data $\mathbf{y}$ by incrementally adding Gaussian noise to it over T discrete timesteps. Let the noisy version of $\mathbf{y}$ at some time step $t \in [0, 1]$ be the *latent variables* $\mathbf{z}_t$, we may define the *variance-preserving* mapping from $\mathbf{y}$ to $\mathbf{z}_t$ (see [233], [460]) as:

$$q(\mathbf{z}_t|\mathbf{y}) = \mathcal{N}\left(\mathbf{z}_t|\alpha_t \mathbf{y}, \sigma_t^2 \mathbf{I}\right), \tag{B.4}$$

where σ_t^2 is a strictly positive scalar-valued function of t that represents the variance of the noise added at each step, and $\alpha_t = \sqrt{1 - \sigma_t^2}$. Note that the distribution $q(\mathbf{z}_t|\mathbf{y})$ is conditioned on the data $\mathbf{y}$.

One can define a *signal-to-noise ratio* SNR(t) $\equiv \alpha_t^2/\sigma_t^2$. As t progresses from 0 (least noisy) to 1 (most noisy), we require that σ_t^2 increases, indicating that more noise is being added, which in turn reduces the signal-to-noise ratio SNR(t). With a sufficiently small SNR(1), the variance-preserving diffusion transformation ensures that the maximally noise-

corrupted data at $t = 1$ follow a standard Gaussian distribution, i.e.

$$q(\mathbf{z}_1|\mathbf{y}) \approx \mathcal{N}(\mathbf{z}_1|0, \mathbf{I}). \tag{B.5}$$

The condition offers key benefits, including analytical tractability, consistency in variance, and ease of sampling, which collectively enhance model stability and efficiency.

As mentioned in Section 4.3.2, the noise schedule, which controls how σ_t^2 depends on t, is the most critical part for the performance of diffusion models [228]. In our framework, we employ a linear noise schedule with the following functional form:

$$\sigma_t^2 = \texttt{Sigmoid}(\gamma_\eta(t)), \quad \text{where } \gamma_\eta(t) = \gamma_{\max} - (\gamma_{\max} - \gamma_{\min})t, \tag{B.6}$$

where $\eta = \{\gamma_{\max}, \gamma_{\min}\}$ is the trainable parameter. It is straightforward to see that SNR(t) = $\exp(-\gamma(t))$, so $\gamma_{\max}$ and $\gamma_{\min}$ also controls the minimum and maximum signal-to-noise ratio of the forward diffusion process. During our training, we initialize $\{\gamma_{\min}, \gamma_{\max}\}$ to be $\{-8, 14\}$, though there is no restriction of what values they can take. Other possibilities for the noise schedule include, for example, a linear or cosine noise schedule with fixed hyperparameters, or using a monotonic neural network (in which case η will be the parameters of the network).

B.2.2 Reverse diffusion

To generate new data, we would like to revert the forward diffusion process in Equation B.4 by gradually denoising the corrupted data at $t = 1$ over a series of discrete timesteps T. This involves training a generative model that can sample a sequence of latent variables $\mathbf{z}_t$ with time moving backward from $t = 1$ to $t = 0$. Define the time steps to be $s(i) = (i-1)/T$ and $t(i) = i/T$ where $i \in [0, T]$, our generative model for data $\mathbf{y}$ can be written as:

$$p(\mathbf{y}) = p(\mathbf{z}_1)p(\mathbf{y}|\mathbf{z}_0)\prod_{i=1}^{T} p(\mathbf{z}_{s(i)}|\mathbf{z}_{t(i)}). \tag{B.7}$$

Familiar readers may recognize that this formalism is similar to a Markovian hierarchical variational autoencoder (VAE) [461], [462], a generalization of VAEs that incorporates multiple levels of latent hierarchies and a Markov chain generative process. Indeed, VDMs can be thought of as Markovian hierarchical VAEs where the latent dimension is the same as the data dimension and the latent encoding process is pre-defined as a linear Gaussian model (see [234]). This interpretation of VDMs is especially useful for understanding its optimization objective.

As stated in Appendix B.2.1, with sufficiently large T and small SNR(1), we expect $q(\mathbf{z}_1)$

to be a standard Gaussian distribution. We thus also model the first term $p(\mathbf{z}_1)$ as a standard Gaussian, i.e

$$p(\mathbf{z}_1) \approx \mathcal{N}(\mathbf{z}_1|0, \mathbf{I}). \tag{B.8}$$

Similarly, as in the VAE analog, the second term represents the reconstructed data likelihood, which we will simply model as:

$$p(\mathbf{y}|\mathbf{z}_0) \approx \mathcal{N}(\mathbf{y}|\mathbf{z}_1, \sigma^2\mathbf{I}). \tag{B.9}$$

Note that σ is a hyperparameter of the data likelihood and *not* related to the noise schedule. It determines how accurately the data is to be reconstructed during the training process, as well as the relative scaling between each term of the optimization objective (see Appendix B.2.3). In our framework, we set $\sigma = 0.001$. Finally, we choose the last term to be:

$$p(\mathbf{z}_s|\mathbf{y}) = q(\mathbf{z}_s|\mathbf{z}_t, \mathbf{y}), \tag{B.10}$$

where we have rewritten the ground truth $q(\mathbf{z}_s|\mathbf{z}_t)$ to be conditioned on the target data $q(\mathbf{z}_s|\mathbf{z}_t, \mathbf{y})$.

B.2.3 Variational optimization objective

To train the VDM, we minimize the negative Evidence Lower Bound (ELBO, [390]),

$$\begin{aligned} -\log p(\mathbf{y}) \leq -\text{ELBO}(\mathbf{y}) = &- \mathbb{E}_{q(\mathbf{z}_1|\mathbf{y})}\left[D_{\text{KL}}\left(q(\mathbf{z}_1|\mathbf{y})\|p(\mathbf{z}_1)\right)\right] && \text{(B.11)} \\ &+ \mathbb{E}_{q(\mathbf{z}_0|\mathbf{y})}\left[\log p(\mathbf{y}|\mathbf{z}_0)\right] && \text{(B.12)} \\ &+ \mathcal{L}_{\text{diff}}(\mathbf{y}), && \text{(B.13)} \end{aligned}$$

where D_{KL} is the Kullback–Leibler (KL) divergence. The first two terms are referred to as the *prior matching loss* and the *reconstruction loss*, paraappendix4:vdmlleling their counterparts in the VAE analog. The prior matching term minimizes the discrepancy between the final latent distribution and the Gaussian prior, while the reconstruction term ensures the accuracy between the data $\mathbf{y}$ and its reconstruction from the latent variable $\mathbf{z}_0$. It is worth noting that in a common setup where the noise schedule is linear, these two terms are ignored since they do not depend on the parameters of the noise prediction model (see [233] for example). However, as our noise schedule (Equation B.6) consists of trainable parameters η, we do not ignore these terms.

Unlike in the VAE analog, the VDM loss includes an additional term, known as the

forward-reverse consistency loss, which can be written as:

$$\mathcal{L}_{\text{diff}}(\mathbf{y}) = \sum_{i=1}^{T} \mathbb{E}_{q(\mathbf{z}_i|\mathbf{y})} D_{\text{KL}} \left[q(\mathbf{z}_{s(i)}|\mathbf{z}_{t(i)}, \mathbf{y}) \| p(\mathbf{z}_{s(i)}|\mathbf{z}_{t(i)}) \right] \tag{B.14}$$

in case of finite T. This term is minimized when the forward diffusion process $q(\mathbf{z}_s|\mathbf{z}_t, \mathbf{y})$ matches the reverse generative process $p_\varphi(\mathbf{z}_s|\mathbf{z}_t)$ across *all time steps*.

In practice, we parameterize the denoising model as a noise prediction model $\hat{\boldsymbol{\epsilon}}_\varphi(\mathbf{z}_t, t)$, which is a neural network with parameters φ. Additionally, the sum over the timesteps in Equation B.14 can be written as an expectation over the timesteps $t(i)$, which is sampled uniformly (i.e., over $i \sim U(1, T)$). We can thus simplify Equation B.14 to be:

$$\mathcal{L}_{\text{diff}}(\mathbf{y}) = \frac{T}{2} \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\mathbf{I}), i \sim U(1,T)} \left[w_\eta(t) \| \boldsymbol{\epsilon} - \hat{\boldsymbol{\epsilon}}_\varphi(\mathbf{z}_t, t) \|_2^2 \right], \tag{B.15}$$

where the pre-factor $w(t)$ can be written as:

$$w_\eta(t) = \exp(\gamma_\eta(t) - \gamma_\eta(s)) - 1. \tag{B.16}$$

A step-by-step derivation of the diffusion loss can be found in [234]. Here, we note that rewriting $\mathcal{L}_{\text{diff}}(\mathbf{y})$ as an expectation over the timesteps $t(i)$ allows for the negative ELBO through a Monte Carlo estimator. This greatly increases the tractability and stability of the optimization process, as we do not need to simulate the entire trajectory from $\mathbf{y}$ to $\mathbf{z}_0$. Equation B.15 optimizes the parameter φ of the noise prediction model and $\eta = \{\gamma_{\max}, \gamma_{\min}\}$ of the noise schedule simultaneously.

Lastly, in case of a conditional generative process, we simply include the conditioning features $\boldsymbol{\theta}$ into the noise prediction model, i.e. $\hat{\boldsymbol{\epsilon}}_\varphi(\mathbf{z}_t, t, \boldsymbol{\theta})$.

B.2.4 Continuous-time diffusion

The number of diffusion steps T is an important hyperparameter of the VDM, with a higher T generally leading to better performance (see Appendix F of [235]). In our framework, we employ a continuous-time VDM, corresponding to setting $T \to \infty$ and also reducing the number of hyperparameters by one. In this limit, the summation over the timesteps in Equation B.14 is simply replaced by an integration over t from $t = 0$ to $t = 1$. The final diffusion loss becomes:

$$\mathcal{L}_{\text{diff}}(\mathbf{y}) = \frac{1}{2} \mathbb{E}_{\epsilon \sim \mathcal{N}(0,\mathbf{I}), t \sim U(0,1)} \left[\gamma'_\eta(t) \| \boldsymbol{\epsilon} - \hat{\boldsymbol{\epsilon}}_\varphi(\mathbf{z}_t, t) \|_2^2 \right], \tag{B.17}$$

where the time derivative $\gamma'_\eta(t) = d\gamma_\eta(t)/dt$ can be evaluated via automatic differentiation. It is worth noting that in this limit, the diffusion process is governed by a stochastic differential equation; see [371] for a more detailed discussion.

B.3 Additional Results

B.3.1 Halo and central subhalo properties

We show additional results for the halo and central subhalo properties recovered by the conditional normalizing flows. As briefly discussed in Section 4.4.1, an important advantage of normalizing flows over traditional regressions is that they can effectively model complex and multi-dimensional distributions.

In Figure B.1, each panel shows the 1σ contours of a few notable parameter combinations. From left to right, the columns show the correlations between (1) the halo mass and central subhalo mass $M_{\rm halo} - M_{\rm sub,c}$, (2) the halo stellar mass and central subhalo stellar mass $M_\star - M_{\star,\rm c}$, (3) the stellar mass and total mass of the central subhalo $M_{\rm sub,c} - M_{\star,\rm c}$, (4) the DM concentration and stellar mass of the central subhalo $c_{\rm sub,c} - M_{\star,\rm c}$, and (5) the DM concentration and total mass of the central subhalo $c_{\rm sub,c} - M_{\rm sub,c}$. The top and bottom panels show the contours for the four e_w bins and four κ_w bins, respectively, with each bin denoted by a different color. The contours of `NeHOD` samples and the simulations are shown as solid and dashed lines. We do not include the $M_{\rm WDM}$ and $\epsilon_{\rm f,high}$ bins as changing these parameters does not significantly alter the contours.

There is a strong correlation between the masses of the halo and the central subhalo, which is expected since the central subhalo dominates the halo mass. We see the 1σ contours predicted by the flows agree well with the simulations for both the e_w and κ_w bins. Similarly, in the third column, the contours of $M_{\rm sub,c} - M_{\star,\rm c}$ also show good agreement between the generated samples and the simulations.

Perhaps most interesting is the relationship between the DM concentration and the stellar mass of the central subhalos in the fourth column. As evidenced, the distribution of the DM concentration is bimodal, suggesting two distinct populations of central subhalos. This bi-modality depends on the stellar mass, which, in turn, is influenced by the stellar feedback parameters e_w and κ_w. Increasing either e_w and κ_w will result in halos with lower stellar mass and DM concentration, which is expected as feedback processes expel gas and reduce star formation efficiency. It is important to note that we do not observe a similar dependency on the subhalo total mass, where the bi-modality occurs at all mass ranges (as shown in the fifth column). This suggests that stellar feedback plays an important role in shaping

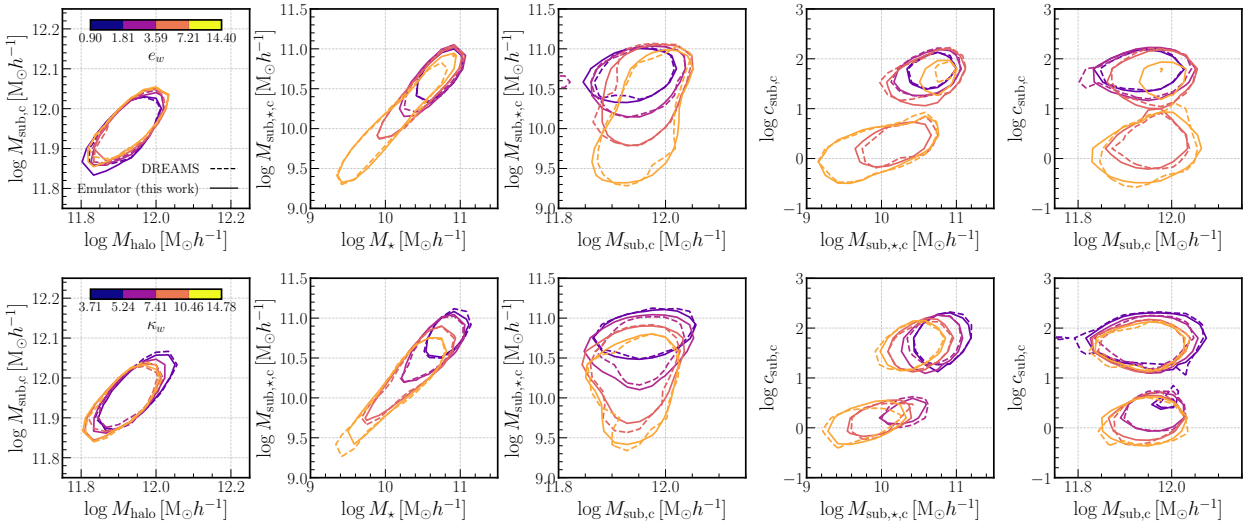

Figure B.1: The 1σ contours of different combinations of halo and central properties. From left to right, the columns show the correlations between (1) the halo mass and central subhalo mass $M_{\rm halo} - M_{\rm sub,c}$, (2) the halo stellar mass and central subhalo stellar mass $M_\star - M_{\star,\rm c}$, (3) the stellar mass and total mass of the central subhalo $M_{\rm sub,c} - M_{\star,\rm c}$, (4) the DM concentration and stellar mass of the central subhalo $c_{\rm sub,c} - M_{\star,\rm c}$, and (5) the DM concentration and total mass of the central subhalo $c_{\rm sub,c} - M_{\rm sub,c}$. The top and bottom panels show the bins of e_w and e_w, respectively.

the potential well and structure of these subhalos. Despite the complex relations between feedback parameters, stellar mass, and DM concentration, we see that the contours predicted by flows agree well with the simulations. There are some minor discrepancies in the highest e_w bin ($e_w \in [7.21, 14.40]$), where the high-$c_{\rm sub,c}$ population predicted by `NeHOD` is a bit more prominent than in the simulations, though this is most likely due to the limited number of training samples.

B.3.2 Mass functions

In Figures B.2 and B.3, we show the normalized stellar and halo mass functions of the generated samples and the DREAMS simulations. This allows us to highlight and better compare the *shape* of the mass functions. Overall, the general trends of both the stellar and halo mass functions of the `NeHOD` samples align well with the simulations. The most significant discrepancy occurs at the highest e_w bin ($e_w \in [7.21, 14.40]$). However, we note that the stellar mass functions of the simulations appear somewhat anomalous in this case. The highest e_w line seems to converge with the second-highest bin, which is unusual. Additionally, we note that the mass functions for the highest and lowest bins are not as accurately recovered as those for the middle bins, though this is most likely due to the edge effects of the prior distributions.

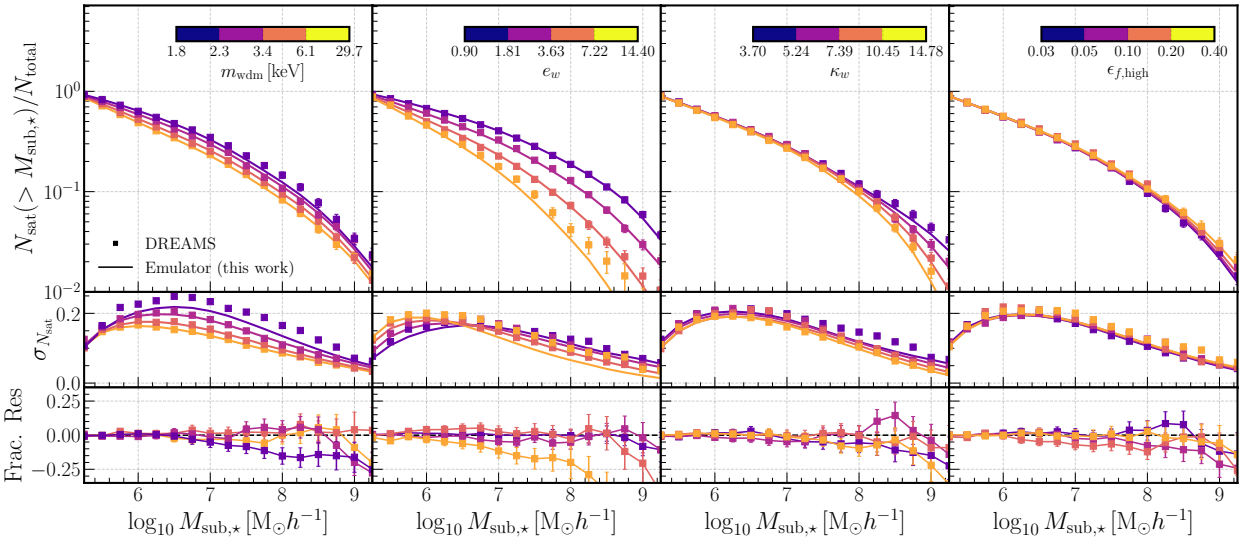

Figure B.2: The normalized satellite stellar mass functions as emulated by `NeHOD` and from the DREAMS simulations. Panels are similar to Figure 4.3.

As also discussed in Section 4.4.2, the VDM slightly underestimates the number of massive subhalos, though it can be seen more clearly with the normalized mass functions. This is likely due to the rarity of these subhalos in the training dataset compared to the rest of the population.

B.3.3 Dark matter concentration

In Figure B.4, we show the 2-D contours for the DM concentration and the total mass of the generated satellites (solid lines) and the simulations (dashed lines) for the same bins of WDM mass and astrophysical parameters. The top panels show the 1σ contours, while the bottom panels show the 2σ contours.

Figure B.4 demonstrates the complex relationship between DM concentration, satellite mass, and the feedback parameters. The distribution of DM concentration is bimodal, though this bi-modality is much less pronounced than observed with the central subhalos (Figure B.1) due to the lower mass range. In general, we see the contours predicted by `NeHOD` align well with those from the simulations. There are minor differences at the high subhalo mass that are also observed in Figure 4.6 in Section 4.4.2. These discrepancies arise because the WDM slightly underestimates the number of high-mass halos, as discussed in Section 4.4.2), due to the limited number of training samples for these subhalos.

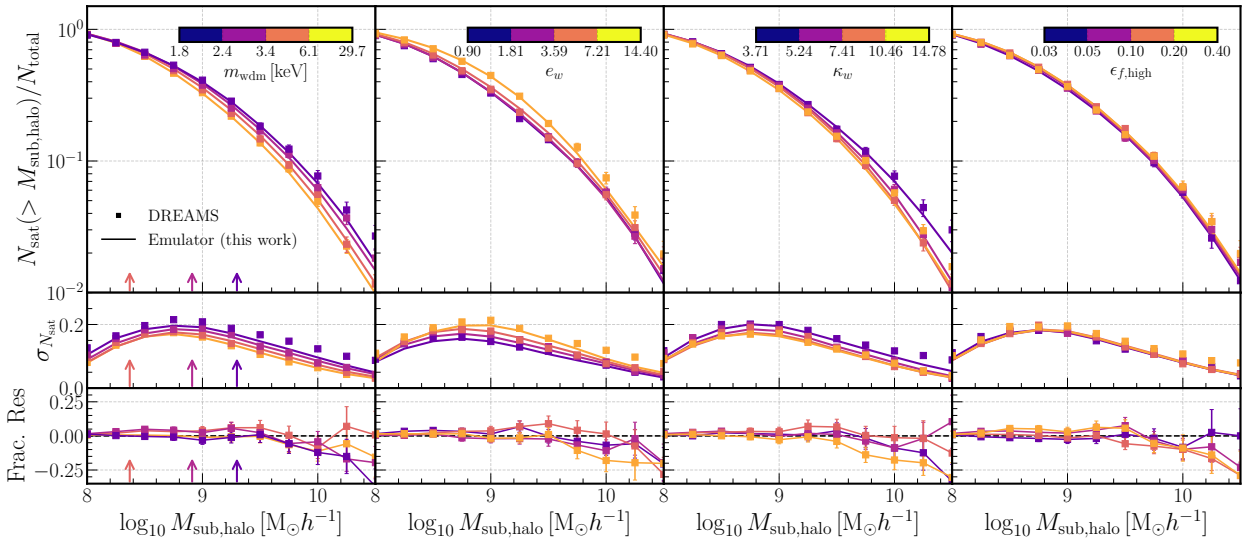

Figure B.3: The normalized satellite halo mass functions as emulated by `NeHOD` and from the DREAMS simulations. Panels are similar to Figure 4.3. The vertical arrows in the first column (WDM mass) show the half-mode mass $M_{\rm hm}$.

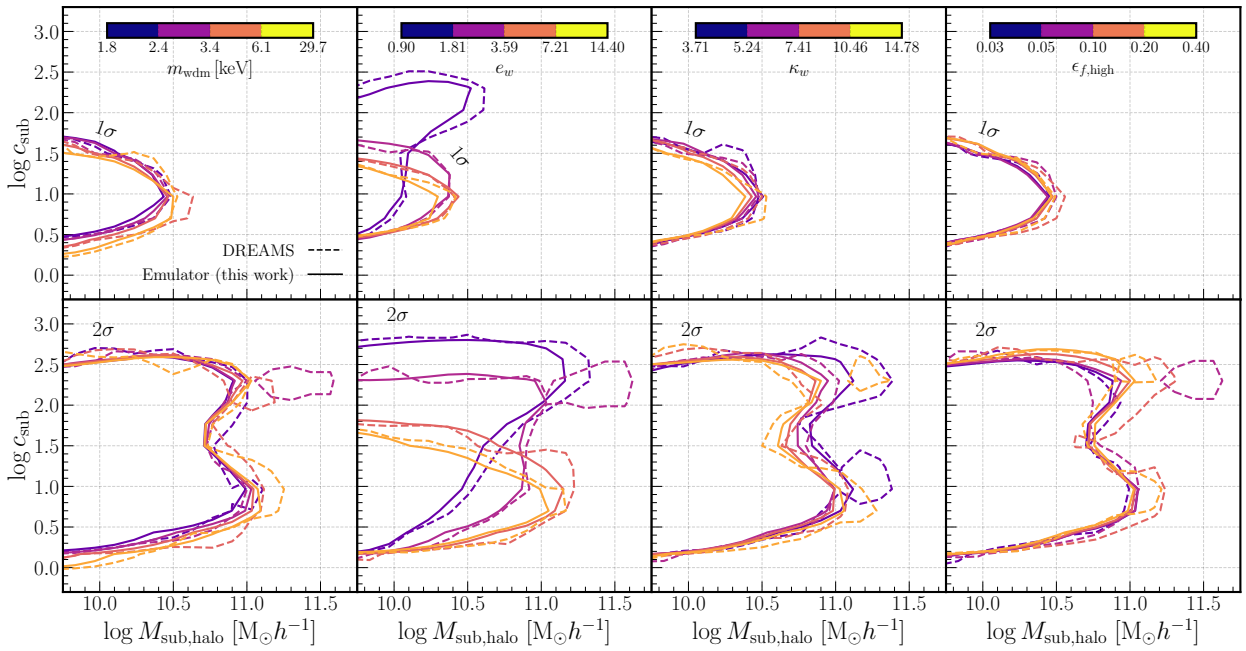

Figure B.4: Contours of the DM concentration and total mass of satellite subhalos generated by `NeHOD` (solid lines) and extracted from the DREAMS simulations (dashed lines). The top and bottom panels show the 1σ and 2σ contours, respectively. Each column shows the contours for different bins of the WDM mass and astrophysical parameters (left to right), with each bin denoted by a different color.

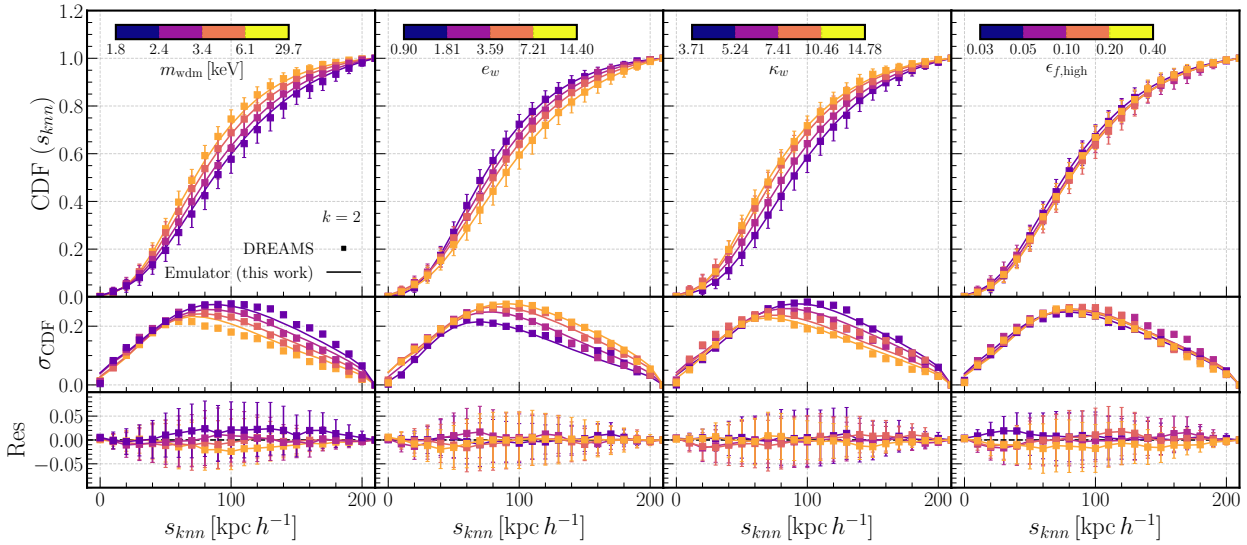

Figure B.5: The CDFs of the k-nearest neighbor distances for $k = 2$. Panels are the same as Figure 4.7.

B.3.4 Position and velocity clustering

We show additional results for the k-nearest neighbor distance and velocity distributions in Section 4.4.2 for $k = 2$ and $k = 4$. The CDFs of the k-nearest neighbor distances are shown in Figure B.5 ($k = 2$) and Figure B.6 ($k = 4$), and velocities are shown in Figure B.7 ($k = 2$) and Figure B.8 ($k = 4$).

In general, the trends are similar to what observed in Figures 4.8 and 4.10, though both the distance and velocity distributions tend to shift towards higher values as the number of nearest neighbor k increases. The performance of the VDM is similar as observed in the $k = 1$ case. The VDM can effectively recover the average CDFs and the standard deviations for the k-nearest neighbor distance. Likewise, we see slight underestimations in the k-nearest neighbor velocity CDFs. This effect is also minimal, at about 5%, and the VDM can capture the general trends of the CDFs accurately.

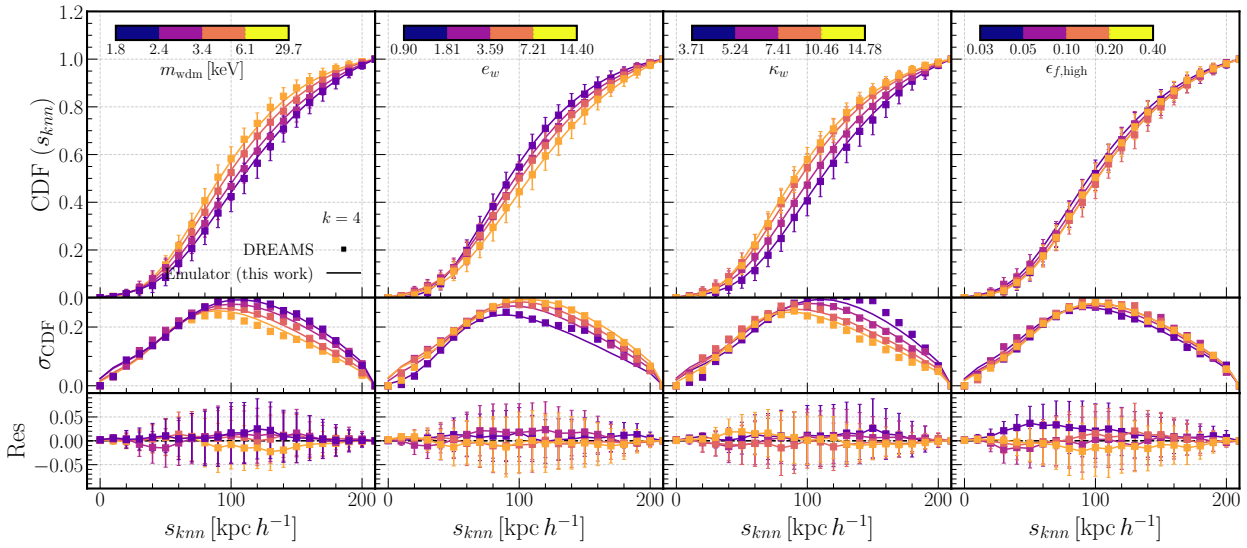

Figure B.6: The CDFs of the k-nearest neighbor distances for $k = 4$. Panels are the same as Figure 4.7.

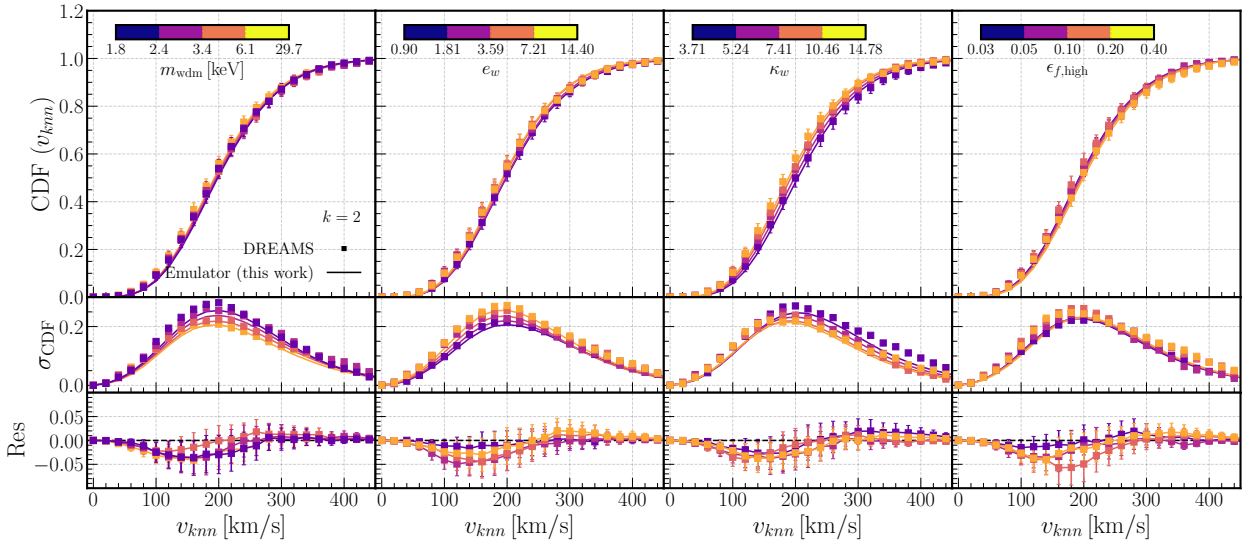

Figure B.7: The CDFs of the k-nearest neighbor velocities for $k = 2$. Panels are the same as Figure 4.7.

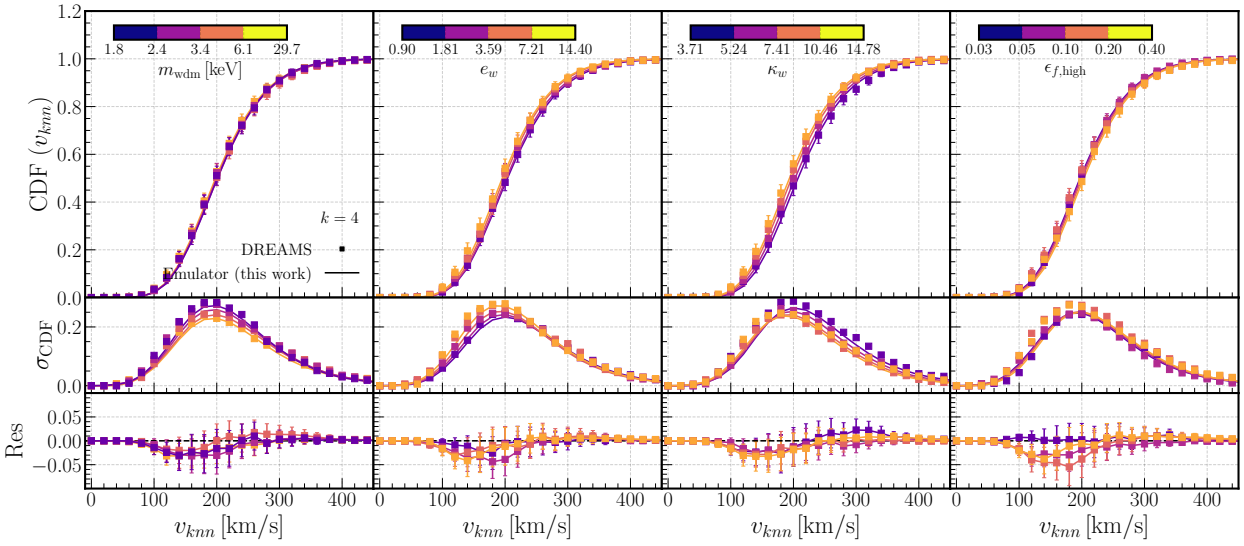

Figure B.8: The CDFs of the k-nearest neighbor velocities for $k = 4$. Panels are the same as Figure 4.7.

References

[1] G. Bertone and D. Hooper, "History of dark matter," *Reviews of Modern Physics*, vol. 90, no. 4, 045002, p. 045 002, Oct. 2018. DOI: 10.1103/RevModPhys.90.045002. arXiv: 1605.04909 [astro-ph.CO].

[2] K. Freese, "Status of dark matter in the universe," *International Journal of Modern Physics D*, vol. 26, no. 6, 1730012-223, pp. 1 730 012–223, Jan. 2017. DOI: 10.1142/S0218271817300129. arXiv: 1701.01840 [astro-ph.CO].

[3] N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," Jul. 2018. arXiv: 1807.06209 [astro-ph.CO].

[4] V. Trimble, "Existence and nature of dark matter in the universe.," *ARA&A*, vol. 25, pp. 425–472, Jan. 1987. DOI: 10.1146/annurev.aa.25.090187.002233.

[5] G. Bertone and T. M. P. Tait, "A new era in the search for dark matter," *Nature*, vol. 562, no. 7725, pp. 51–56, Oct. 2018. DOI: 10.1038/s41586-018-0542-z. arXiv: 1810.01668 [astro-ph.CO].

[6] Y. Sofue and V. Rubin, "Rotation Curves of Spiral Galaxies," *ARA&A*, vol. 39, pp. 137–174, Jan. 2001. DOI: 10.1146/annurev.astro.39.1.137. arXiv: astro-ph/0010594 [astro-ph].

[7] V. C. Rubin, J. Ford W. K., and N. Thonnard, "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc).," *ApJ*, vol. 238, pp. 471–487, Jun. 1980. DOI: 10.1086/158003.

[8] Y. Sofue, M. Honma, and T. Omodaka, "Unified Rotation Curve of the Galaxy – Decomposition into de Vaucouleurs Bulge, Disk, Dark Halo, and the 9-kpc Rotation Dip –," *PASJ*, vol. 61, p. 227, Feb. 2009. DOI: 10.1093/pasj/61.2.227. arXiv: 0811.0859 [astro-ph].

[9] C. Wegg, O. Gerhard, and M. Bieth, "The gravitational force field of the Galaxy measured from the kinematics of RR Lyrae in Gaia," *MNRAS*, vol. 485, no. 3, pp. 3296–3316, May 2019. DOI: 10.1093/mnras/stz572. arXiv: 1806.09635 [astro-ph.GA].

[10] P. R. Kafle, S. Sharma, G. F. Lewis, and J. Bland-Hawthorn, "Kinematics of the Stellar Halo and the Mass Distribution of the Milky Way Using Blue Horizontal Branch Stars," *ApJ*, vol. 761, no. 2, 98, p. 98, Dec. 2012. DOI: 10.1088/0004-637X/761/2/98. arXiv: 1210.7527 [astro-ph.GA].

[11] A. V. Zasov, A. S. Saburova, A. V. Khoperskov, and S. A. Khoperskov, "Dark matter in galaxies," *Physics Uspekhi*, vol. 60, no. 1, p. 3, Apr. 2017. DOI: 10.3367/UFNe.2016.03.037751. arXiv: 1710.10630 [astro-ph.GA].

[12] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln," *Helvetica Physica Acta*, vol. 6, pp. 110–127, Jan. 1933.

[13] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A Direct Empirical Proof of the Existence of Dark Matter," *ApJ*, vol. 648, no. 2, pp. L109–L113, Sep. 2006. DOI: 10.1086/508162. arXiv: astro-ph/0608407 [astro-ph].

[14] R. Wojtak and E. L. Łokas, "Mass profiles and galaxy orbits in nearby galaxy clusters from the analysis of the projected phase space," *MNRAS*, vol. 408, no. 4, pp. 2442–2456, Nov. 2010. DOI: 10.1111/j.1365-2966.2010.17297.x. arXiv: 1004.3771 [astro-ph.CO].

[15] C. Pryor and J. Kormendy, "The Dark Matter Halos of Draco and Ursa Minor," *AJ*, vol. 100, p. 127, Jul. 1990. DOI: 10.1086/115496.

[16] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi, "Dark Matter Substructure within Galactic Halos," *ApJ*, vol. 524, no. 1, pp. L19–L22, Oct. 1999. DOI: 10.1086/312287. arXiv: astro-ph/9907411 [astro-ph].

[17] L. E. Strigari, J. S. Bullock, M. Kaplinghat, J. D. Simon, M. Geha, B. Willman, and M. G. Walker, "A common mass scale for satellite galaxies of the Milky Way," *Nature*, vol. 454, no. 7208, pp. 1096–1097, Aug. 2008. DOI: 10.1038/nature07222. arXiv: 0808.3772 [astro-ph].

[18] M. G. Walker, M. Mateo, E. W. Olszewski, J. Peñarrubia, N. W. Evans, and G. Gilmore, "A Universal Mass Profile for Dwarf Spheroidal Galaxies?" *ApJ*, vol. 704, no. 2, pp. 1274–1287, Oct. 2009. DOI: 10.1088/0004-637X/704/2/1274. arXiv: 0906.0341 [astro-ph.CO].

[19] R. S. Ellis, "Gravitational lensing: a unique probe of dark matter and dark energy," *Philosophical Transactions of the Royal Society of London Series A*, vol. 368, no. 1914, pp. 967–987, Feb. 2010. DOI: 10.1098/rsta.2009.0209.

[20] *Gravitational Lensing: Strong, Weak and Micro*, Jan. 2006. DOI: 10.48550/arXiv.astro-ph/0407232. arXiv: astro-ph/0407232 [astro-ph].

[21] R. Massey, T. Kitching, and J. Richard, "The dark matter of gravitational lensing," *Reports on Progress in Physics*, vol. 73, no. 8, 086901, p. 086 901, Aug. 2010. DOI: 10.1088/0034-4885/73/8/086901. arXiv: 1001.1739 [astro-ph.CO].

[22] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, W. Forman, C. Jones, S. Murray, and W. Tucker, "Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56," *ApJ*, vol. 606, no. 2, pp. 819–824, May 2004. DOI: 10.1086/383178. arXiv: astro-ph/0309303 [astro-ph].

[23] M. Bradač, D. Clowe, A. H. Gonzalez, P. Marshall, W. Forman, C. Jones, M. Markevitch, S. Randall, T. Schrabback, and D. Zaritsky, "Strong and Weak Lensing United. III. Measuring the Mass Distribution of the Merging Galaxy Cluster 1ES 0657-558," *ApJ*, vol. 652, no. 2, pp. 937–947, Dec. 2006. DOI: 10.1086/508601. arXiv: astro-ph/0608408 [astro-ph].

[24] M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis.," *ApJ*, vol. 270, pp. 365–370, Jul. 1983. DOI: 10.1086/161130.

[25] T. H. Randriamampandry and C. Carignan, "Galaxy mass models: MOND versus dark matter haloes," *MNRAS*, vol. 439, no. 2, pp. 2132–2145, Apr. 2014. DOI: 10.1093/mnras/stu100. arXiv: 1401.5619 [astro-ph.GA].

[26] J. R. Brownstein and J. W. Moffat, "The Bullet Cluster 1E0657-558 evidence shows modified gravity in the absence of dark matter," *MNRAS*, vol. 382, no. 1, pp. 29–47, Nov. 2007. DOI: 10.1111/j.1365-2966.2007.12275.x. arXiv: astro-ph/0702146 [astro-ph].

[27] G. F. Smoot, C. L. Bennett, A. Kogut, *et al.*, "Structure in the COBE Differential Microwave Radiometer First-Year Maps," *ApJ*, vol. 396, p. L1, Sep. 1992. DOI: 10.1086/186504.

[28] G. Hinshaw, D. Larson, E. Komatsu, *et al.*, "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *ApJS*, vol. 208, no. 2, 19, p. 19, Oct. 2013. DOI: 10.1088/0067-0049/208/2/19. arXiv: 1212.5226 [astro-ph.CO].

[29] C. L. Bennett, M. Halpern, G. Hinshaw, *et al.*, "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results," *ApJS*, vol. 148, no. 1, pp. 1–27, Sep. 2003. DOI: 10.1086/377253. arXiv: astro-ph/0302207 [astro-ph].

[30] G. Hinshaw, D. N. Spergel, L. Verde, *et al.*, "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: The Angular Power Spectrum," *ApJS*, vol. 148, no. 1, pp. 135–159, Sep. 2003. DOI: 10.1086/377225. arXiv: astro-ph/0302217 [astro-ph].

[31] Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters," *A&A*, vol. 571, A16, A16, Nov. 2014. DOI: 10.1051/0004-6361/201321591. arXiv: 1303.5076 [astro-ph.CO].

[32] Planck Collaboration, N. Aghanim, Y. Akrami, *et al.*, "Planck 2018 results. VI. Cosmological parameters," *A&A*, vol. 641, A6, A6, Sep. 2020. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[33] D. G. York, J. Adelman, J. Anderson John E., *et al.*, "The Sloan Digital Sky Survey: Technical Summary," *AJ*, vol. 120, no. 3, pp. 1579–1587, Sep. 2000. DOI: 10.1086/301513. arXiv: astro-ph/0006396 [astro-ph].

[34] V. Springel, S. D. M. White, A. Jenkins, *et al.*, "Simulations of the formation, evolution and clustering of galaxies and quasars," *Nature*, vol. 435, no. 7042, pp. 629–636, Jun. 2005. DOI: 10.1038/nature03597. arXiv: astro-ph/0504097 [astro-ph].

[35] A. Ragagnin, M. Meneghetti, F. Calura, G. Despali, K. Dolag, M. S. Fischer, C. Giocoli, and L. Moscardini, "Dianoga SIDM: galaxy cluster self-interacting dark matter simulations," *arXiv e-prints*, arXiv:2404.01383, arXiv:2404.01383, Apr. 2024. DOI: 10.48550/arXiv.2404.01383. arXiv: 2404.01383 [astro-ph.CO].

[36] S. Tulin and H.-B. Yu, "Dark Matter Self-interactions and Small Scale Structure," *Phys. Rept.*, vol. 730, pp. 1–57, 2018. DOI: 10.1016/j.physrep.2017.11.004. arXiv: 1705.02358 [hep-ph].

[37] E. O. Nadler, A. Drlica-Wagner, K. Bechtol, *et al.*, "Constraints on Dark Matter Properties from Observations of Milky Way Satellite Galaxies," *Phys. Rev. Lett.*, vol. 126, no. 9, 091101, p. 091 101, Mar. 2021. DOI: 10.1103/PhysRevLett.126.091101. arXiv: 2008.00022 [astro-ph.CO].

[38] J. Zavala, M. Vogelsberger, and M. G. Walker, "Constraining self-interacting dark matter with the Milky way's dwarf spheroidals.," *MNRAS*, vol. 431, pp. L20–L24, Apr. 2013. DOI: 10.1093/mnrasl/sls053. arXiv: 1211.6426 [astro-ph.CO].

[39] S. D. M. White and M. J. Rees, "Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering.," *MNRAS*, vol. 183, pp. 341–358, May 1978. DOI: 10.1093/mnras/183.3.341.

[40] A. Pillepich, V. Springel, D. Nelson, *et al.*, "Simulating galaxy formation with the IllustrisTNG model," *MNRAS*, vol. 473, no. 3, pp. 4077–4106, Jan. 2018. DOI: 10.1093/mnras/stx2656. arXiv: 1703.02970 [astro-ph.GA].

[41] R. Weinberger, V. Springel, L. Hernquist, *et al.*, "Simulating galaxy formation with black hole driven thermal and kinetic feedback," *MNRAS*, vol. 465, no. 3, pp. 3291–3308, Mar. 2017. DOI: 10.1093/mnras/stw2944. arXiv: 1607.03486 [astro-ph.GA].

[42] R. A. Crain, J. Schaye, R. G. Bower, *et al.*, "The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations," *MNRAS*, vol. 450, no. 2, pp. 1937–1961, Jun. 2015. DOI: 10.1093/mnras/stv725. arXiv: 1501.01311 [astro-ph.GA].

[43] P. F. Hopkins, D. Kereš, J. Oñorbe, C.-A. Faucher-Giguère, E. Quataert, N. Murray, and J. S. Bullock, "Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation," *MNRAS*, vol. 445, no. 1, pp. 581–603, Nov. 2014. DOI: 10.1093/mnras/stu1738. arXiv: 1311.2073 [astro-ph.CO].

[44] C. S. Frenk and S. D. M. White, "Dark matter and cosmic structure," *Annalen der Physik*, vol. 524, no. 9-10, pp. 507–534, Oct. 2012. DOI: 10.1002/andp.201200212. arXiv: 1210.0544 [astro-ph.CO].

[45] J. S. Bullock and M. Boylan-Kolchin, "Small-Scale Challenges to the ΛCDM Paradigm," *Ann. Rev. Astron. Astrophys.*, vol. 55, pp. 343–387, 2017. DOI: 10.1146/annurev-astro-091916-055313. arXiv: 1707.04256 [astro-ph.CO].

[46] P. Bode, J. P. Ostriker, and N. Turok, "Halo Formation in Warm Dark Matter Models," *ApJ*, vol. 556, no. 1, pp. 93–107, Jul. 2001. DOI: 10.1086/321541. arXiv: astro-ph/0010389 [astro-ph].

[47] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto, "Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-α forest," *Phys. Rev. D*, vol. 71, no. 6, 063534, p. 063 534, Mar. 2005. DOI: 10.1103/PhysRevD.71.063534. arXiv: astro-ph/0501562 [astro-ph].

[48] D. N. Spergel and P. J. Steinhardt, "Observational Evidence for Self-Interacting Cold Dark Matter," *Phys. Rev. Lett.*, vol. 84, no. 17, pp. 3760–3763, Apr. 2000. DOI: 10.1103/PhysRevLett.84.3760. arXiv: astro-ph/9909386 [astro-ph].

[49] V. H. Robles, J. S. Bullock, O. D. Elbert, A. Fitts, A. González-Samaniego, M. Boylan-Kolchin, P. F. Hopkins, C.-A. Faucher-Giguère, D. Kereš, and C. C. Hayward, "SIDM on fire: hydrodynamical self-interacting dark matter simulations of low-mass dwarf galaxies," *MNRAS*, vol. 472, pp. 2945–2954, Dec. 2017. DOI: 10.1093/mnras/stx2253. arXiv: 1706.07514.

[50] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, *et al.*, "The Gaia mission," *A&A*, vol. 595, A1, A1, Nov. 2016. DOI: 10.1051/0004-6361/201629272. arXiv: 1609.04153 [astro-ph.IM].

[51] K. Malhan and R. A. Ibata, "Constraining the Milky Way halo potential with the GD-1 stellar stream," *MNRAS*, vol. 486, no. 3, pp. 2995–3005, Jul. 2019. DOI: 10.1093/mnras/stz1035. arXiv: 1807.05994 [astro-ph.GA].

[52] X. Ou, A.-C. Eilers, L. Necib, and A. Frebel, "The dark matter profile of the Milky Way inferred from its circular velocity curve," *arXiv e-prints*, arXiv:2303.12838, arXiv:2303.12838, Mar. 2023. DOI: 10.48550/arXiv.2303.12838. arXiv: 2303.12838 [astro-ph.GA].

[53] C. Roche, L. Necib, T. Lin, X. Ou, and T. Nguyen, "The Escape Velocity Profile of the Milky Way from Gaia DR3," *arXiv e-prints*, arXiv:2402.00108, arXiv:2402.00108, Jan. 2024. DOI: 10.48550/arXiv.2402.00108. arXiv: 2402.00108 [astro-ph.GA].

[54] L. Necib, M. Lisanti, S. Garrison-Kimmel, A. Wetzel, R. Sanderson, P. F. Hopkins, C.-A. Faucher-Giguère, and D. Kereš, "Under the FIRElight: Stellar Tracers of the Local Dark Matter Velocity Distribution in the Milky Way," *ApJ*, vol. 883, no. 1, 27, p. 27, Sep. 2019. DOI: 10.3847/1538-4357/ab3afc. arXiv: 1810.12301 [astro-ph.GA].

[55] E. Aprile, J. Aalbers, F. Agostini, *et al.*, "First Dark Matter Search Results from the XENON1T Experiment," *Phys. Rev. Lett.*, vol. 119, no. 18, 181301, p. 181 301, Nov. 2017. DOI: 10.1103/PhysRevLett.119.181301. arXiv: 1705.06655 [astro-ph.CO].

[56] D. S. Akerib, C. W. Akerlof, S. K. Alsum, *et al.*, "Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment," *Phys. Rev. D*, vol. 101, no. 5, 052002, p. 052 002, Mar. 2020. DOI: 10.1103/PhysRevD.101.052002. arXiv: 1802.06039 [astro-ph.IM].

[57] N. Shipp, D. Erkal, A. Drlica-Wagner, *et al.*, "Measuring the Mass of the Large Magellanic Cloud with Stellar Streams Observed by S ^{5}," *ApJ*, vol. 923, no. 2, 149, p. 149, Dec. 2021. DOI: 10.3847/1538-4357/ac2e93. arXiv: 2107.13004 [astro-ph.GA].

[58] T. Hilmi, D. Erkal, S. E. Koposov, *et al.*, "Inferring dark matter subhalo properties from simulated subhalo-stream encounters," *arXiv e-prints*, arXiv:2404.02953, arXiv:2404.02953, Apr. 2024. DOI: 10.48550/arXiv.2404.02953. arXiv: 2404.02953 [astro-ph.GA].

[59] V. Belokurov, D. B. Zucker, N. W. Evans, *et al.*, "The Field of Streams: Sagittarius and Its Siblings," *ApJ*, vol. 642, no. 2, pp. L137–L140, May 2006. DOI: 10.1086/504797. arXiv: astro-ph/0605025 [astro-ph].

[60] N. Shipp, A. Drlica-Wagner, E. Balbinot, *et al.*, "Stellar Streams Discovered in the Dark Energy Survey," *ApJ*, vol. 862, no. 2, 114, p. 114, Aug. 2018. DOI: 10.3847/1538-4357/aacdab. arXiv: 1801.03097 [astro-ph.GA].

[61] T. S. Li, S. E. Koposov, D. B. Zucker, *et al.*, "The southern stellar stream spectroscopic survey (S^5): Overview, target selection, data reduction, validation, and early science," *MNRAS*, vol. 490, no. 3, pp. 3508–3531, Dec. 2019. DOI: 10.1093/mnras/stz2731. arXiv: 1907.09481 [astro-ph.GA].

[62] DES Collaboration, "The Dark Energy Survey: more than dark energy - an overview," *MNRAS*, vol. 460, pp. 1270–1299, Aug. 2016. DOI: 10.1093/mnras/stw641. arXiv: 1601.00329.

[63] P. Bode, J. P. Ostriker, and N. Turok, "Halo Formation in Warm Dark Matter Models," *ApJ*, vol. 556, no. 1, pp. 93–107, Jul. 2001. DOI: 10.1086/321541. arXiv: astro-ph/0010389 [astro-ph].

[64] S. R. Majewski, R. P. Schiavon, P. M. Frinchaboy, *et al.*, "The Apache Point Observatory Galactic Evolution Experiment (APOGEE)," *AJ*, vol. 154, no. 3, 94, p. 94, Sep. 2017. DOI: 10.3847/1538-3881/aa784d. arXiv: 1509.05420 [astro-ph.IM].

[65] B. Ostdiek, L. Necib, T. Cohen, M. Freytsis, M. Lisanti, S. Garrison-Kimmmel, A. Wetzel, R. E. Sanderson, and P. F. Hopkins, "Cataloging accreted stars within Gaia DR2 using deep learning," *A&A*, vol. 636, A75, A75, Apr. 2020. DOI: 10.1051/0004-6361/201936866. arXiv: 1907.06652 [astro-ph.GA].

[66] L. Necib, B. Ostdiek, M. Lisanti, T. Cohen, M. Freytsis, and S. Garrison-Kimmel, "Chasing Accreted Structures within Gaia DR2 Using Deep Learning," *ApJ*, vol. 903, no. 1, 25, p. 25, Nov. 2020. DOI: 10.3847/1538-4357/abb814. arXiv: 1907.07681 [astro-ph.GA].

[67] L. Necib, B. Ostdiek, M. Lisanti, T. Cohen, M. Freytsis, S. Garrison-Kimmel, P. F. Hopkins, A. Wetzel, and R. Sanderson, "Evidence for a vast prograde stellar stream in the solar vicinity," *Nature Astronomy*, vol. 4, pp. 1078–1083, Jul. 2020. DOI: 10.1038/s41550-020-1131-2. arXiv: 1907.07190 [astro-ph.GA].

[68] S. E. Koposov, M. G. Walker, V. Belokurov, *et al.*, "Snake in the Clouds: a new nearby dwarf galaxy in the Magellanic bridge*," *MNRAS*, vol. 479, no. 4, pp. 5343–5361, Oct. 2018. DOI: 10.1093/mnras/sty1772. arXiv: 1804.06430 [astro-ph.GA].

[69] V. Belokurov, D. Erkal, N. W. Evans, S. E. Koposov, and A. J. Deason, "Co-formation of the disc and the stellar halo," *MNRAS*, vol. 478, no. 1, pp. 611–619, Jul. 2018. DOI: 10.1093/mnras/sty982. arXiv: 1802.03414 [astro-ph.GA].

[70] O. Fakhouri and C.-P. Ma, "The nearly universal merger rate of dark matter haloes in ΛCDM cosmology," *MNRAS*, vol. 386, no. 2, pp. 577–592, May 2008. DOI: 10.1111/j.1365-2966.2008.13075.x. arXiv: 0710.4567 [astro-ph].

[71] C. J. Conselice, A. F. L. Bluck, A. Mortlock, D. Palamara, and A. J. Benson, "Galaxy formation as a cosmological tool - I. The galaxy merger history as a measure of cosmological parameters," *MNRAS*, vol. 444, no. 2, pp. 1125–1143, Oct. 2014. DOI: 10.1093/mnras/stu1385. arXiv: 1407.3811 [astro-ph.GA].

[72] J. F. Navarro, V. R. Eke, and C. S. Frenk, "The cores of dwarf galaxy haloes," *MNRAS*, vol. 283, no. 3, pp. L72–L78, Dec. 1996, ISSN: 0035-8711. DOI: 10.1093/mnras/283.3.L72. eprint: https://academic.oup.com/mnras/article-pdf/283/3/L72/3305901/283-3-L72.pdf. URL: https://doi.org/10.1093/mnras/283.3.L72.

[73] J. I. Read and G. Gilmore, "Mass loss from dwarf spheroidal galaxies: the origins of shallow dark matter cores and exponential surface brightness profiles," *MNRAS*, vol. 356, no. 1, pp. 107–124, Jan. 2005. DOI: 10.1111/j.1365-2966.2004.08424.x. arXiv: astro-ph/0409565 [astro-ph].

[74] S. Mashchenko, H. M. P. Couchman, and J. Wadsley, "The removal of cusps from galaxy centres by stellar feedback in the early Universe," *Nature*, vol. 442, no. 7102, pp. 539–542, Aug. 2006. DOI: 10.1038/nature04944. arXiv: astro-ph/0605672 [astro-ph].

[75] A. Pontzen and F. Governato, "How supernova feedback turns dark matter cusps into cores," *MNRAS*, vol. 421, no. 4, pp. 3464–3471, Apr. 2012. DOI: 10.1111/j.1365-2966.2012.20571.x. arXiv: 1106.0499 [astro-ph.CO].

[76] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky, and O. Ruchayskiy, "The haloes of bright satellite galaxies in a warm dark matter universe," *MNRAS*, vol. 420, no. 3, pp. 2318–2324, Feb. 2012, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.20200.x. eprint: https://academic.oup.com/mnras/article-pdf/420/3/2318/3020178/mnras0420-2318.pdf. URL: https://doi.org/10.1111/j.1365-2966.2011.20200.x.

[77] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, "Core formation in dwarf haloes with self-interacting dark matter: no fine-tuning necessary," *MNRAS*, vol. 453, no. 1, pp. 29–37, Aug. 2015, ISSN: 0035-8711. DOI: 10.1093/mnras/stv1470. eprint: https://academic.oup.com/mnras/article-pdf/453/1/29/4900820/stv1470.pdf. URL: https://doi.org/10.1093/mnras/stv1470.

[78] J. D. Burger, J. Zavala, L. V. Sales, M. Vogelsberger, F. Marinacci, and P. Torrey, "Degeneracies between self-interacting dark matter and supernova feedback as cusp-core transformation mechanisms," *MNRAS*, vol. 513, no. 3, pp. 3458–3481, Jul. 2022. DOI: 10.1093/mnras/stac994. arXiv: 2108.07358 [astro-ph.GA].

[79] J. S. Bullock and M. Boylan-Kolchin, "Small-Scale Challenges to the ΛCDM Paradigm," *ARA&A*, vol. 55, no. 1, pp. 343–387, Aug. 2017. DOI: 10.1146/annurev-astro-091916-055313. arXiv: 1707.04256 [astro-ph.CO].

[80] Y.-Y. Mao, M. Geha, R. H. Wechsler, *et al.*, "The SAGA Survey. III. A Census of 101 Satellite Systems around Milky Way-mass Galaxies," *arXiv e-prints*, arXiv:2404.14498, arXiv:2404.14498, Apr. 2024. DOI: 10.48550/arXiv.2404.14498. arXiv: 2404.14498 [astro-ph.GA].

[81] M. Geha, Y.-Y. Mao, R. H. Wechsler, *et al.*, "The SAGA Survey. IV. The Star Formation Properties of 101 Satellite Systems around Milky Way-mass Galaxies," *arXiv e-prints*, arXiv:2404.14499, arXiv:2404.14499, Apr. 2024. DOI: 10.48550/arXiv.2404.14499. arXiv: 2404.14499 [astro-ph.GA].

[82] Y. Wang, E. O. Nadler, Y.-Y. Mao, *et al.*, "The SAGA Survey. V. Modeling Satellite Systems around Milky Way-mass Galaxies with Updated UniverseMachine," *arXiv e-prints*, arXiv:2404.14500, arXiv:2404.14500, Apr. 2024. DOI: 10.48550/arXiv.2404.14500. arXiv: 2404.14500 [astro-ph.GA].

[83] J. D. Simon, "The Faintest Dwarf Galaxies," *arXiv e-prints*, arXiv:1901.05465, arXiv:1901.05465, Jan. 2019. arXiv: 1901.05465 [astro-ph.GA].

[84] Y. Revaz and P. Jablonka, "Pushing back the limits: detailed properties of dwarf galaxies in a ΛCDM universe," *A&A*, vol. 616, A96, A96, Aug. 2018. DOI: 10.1051/0004-6361/201832669. arXiv: 1801.06222 [astro-ph.GA].

[85] K. El-Badry, A. R. Wetzel, M. Geha, E. Quataert, P. F. Hopkins, D. Kereš, T. K. Chan, and C.-A. Faucher-Giguère, "When the Jeans Do Not Fit: How Stellar Feedback Drives Stellar Kinematics and Complicates Dynamical Modeling in Low-mass Galaxies," *ApJ*, vol. 835, no. 2, 193, p. 193, Feb. 2017. DOI: 10.3847/1538-4357/835/2/193. arXiv: 1610.04232 [astro-ph.GA].

[86] A. Lazar, J. S. Bullock, M. Boylan-Kolchin, *et al.*, "A dark matter profile to model diverse feedback-induced core sizes of ΛCDM haloes," *arXiv e-prints*, arXiv:2004.10817, arXiv:2004.10817, Apr. 2020. arXiv: 2004.10817 [astro-ph.GA].

[87] A. Geringer-Sameth, M. G. Walker, S. M. Koushiappas, S. E. Koposov, V. Belokurov, G. Torrealba, and N. Evans, "Indication of Gamma-ray Emission from the Newly Discovered Dwarf Galaxy Reticulum II," *Phys. Rev. Lett.*, vol. 115, no. 8, p. 081 101, 2015. DOI: 10.1103/PhysRevLett.115.081101. arXiv: 1503.02320 [astro-ph.HE].

[88] A. Geringer-Sameth, S. M. Koushiappas, and M. Walker, "Dwarf galaxy annihilation and decay emission profiles for dark matter experiments," *ApJ*, vol. 801, no. 2, p. 74, 2015. DOI: 10.1088/0004-637X/801/2/74. arXiv: 1408.0002 [astro-ph.CO].

[89] M. G. Walker, M. Mateo, and E. W. Olszewski, "Stellar Velocities in the Carina, Fornax, Sculptor, and Sextans dSph Galaxies: Data From the Magellan/MMFS Survey," *ApJ*, vol. 137, pp. 3100–3108, Feb. 2009. DOI: 10.1088/0004-6256/137/2/3100. arXiv: 0811.0118 [astro-ph].

[90] M. Mateo, E. W. Olszewski, and M. G. Walker, "The Velocity Dispersion Profile of the Remote Dwarf Spheroidal Galaxy Leo I: A Tidal Hit and Run?" *ApJ*, vol. 675, no. 1, pp. 201–233, Mar. 2008. DOI: 10.1086/522326. arXiv: 0708.1327 [astro-ph].

[91] V. Bonnivard, C. Combet, D. Maurin, and M. G. Walker, "Spherical Jeans analysis for dark matter indirect detection in dwarf spheroidal galaxies - impact of physical parameters and triaxiality," *MNRAS*, vol. 446, pp. 3002–3021, Jan. 2015. DOI: 10.1093/mnras/stu2296. arXiv: 1407.7822 [astro-ph.HE].

[92] J. I. Read and P. Steger, "How to break the density-anisotropy degeneracy in spherical stellar systems," *MNRAS*, vol. 471, pp. 4541–4558, Nov. 2017. DOI: 10.1093/mnras/stx1798.

[93] T. Nguyen, S. Mishra-Sharma, R. Williams, and L. Necib, "Uncovering dark matter density profiles in dwarf galaxies with graph neural networks," *Phys. Rev. D*, vol. 107, no. 4, 043015, p. 043 015, Feb. 2023. DOI: 10.1103/PhysRevD.107.043015. arXiv: 2208.12825 [astro-ph.CO].

[94] M. Cirelli, G. Corcella, A. Hektor, G. Hütsi, M. Kadastik, P. Panci, M. Raidal, F. Sala, and A. Strumia, "PPPC 4 DM ID: A poor particle physicist cookbook for dark matter indirect detection," *Journal of Cosmology and Astroparticle Physics*, vol. 2011, no. 03, pp. 051–051, Mar. 2011. DOI: 10.1088/1475-7516/2011/03/051. URL: https://doi.org/10.1088/1475-7516/2011/03/051.

[95] M. Ackermann *et al.*, "Searching for dark matter annihilation from milky way dwarf spheroidal galaxies with six years of fermi large area telescope data," *Phys. Rev. Lett.*, vol. 115, p. 231 301, 23 Nov. 2015. DOI: 10.1103/PhysRevLett.115.231301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.231301.

[96] G. D. Martinez, "A robust determination of Milky Way satellite properties using hierarchical mass modelling," *Monthly Notices of the Royal Astronomical Society*, vol. 451, no. 3, pp. 2524–2535, Jun. 2015, ISSN: 0035-8711. DOI: 10.1093/mnras/stv942. eprint: https://academic.oup.com/mnras/article-pdf/451/3/2524/4032739/stv942.pdf. URL: https://doi.org/10.1093/mnras/stv942.

[97] A. Geringer-Sameth and S. M. Koushiappas, "Exclusion of canonical weakly interacting massive particles by joint analysis of milky way dwarf galaxies with data from the fermi gamma-ray space telescope," *Phys. Rev. Lett.*, vol. 107, p. 241 303, 24 Dec. 2011. DOI: 10.1103/PhysRevLett.107.241303. URL: https://link.aps.org/doi/10.1103/PhysRevLett.107.241303.

[98] M. Ackermann *et al.*, "Constraining dark matter models from a combined analysis of milky way satellites with the fermi large area telescope," *Phys. Rev. Lett.*, vol. 107, p. 241 302, 24 Dec. 2011. DOI: 10.1103/PhysRevLett.107.241302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.107.241302.

[99] M. Mazziotta, F. Loparco, F. de Palma, and N. Giglietto, "A model-independent analysis of the fermi large area telescope gamma-ray data from the milky way dwarf galaxies and halo to constrain dark matter scenarios," *Astroparticle Physics*, vol. 37, pp. 26–39, 2012, ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.07.005. URL: https://www.sciencedirect.com/science/article/pii/S0927650512001417.

[100] A. Geringer-Sameth, S. M. Koushiappas, and M. G. Walker, "Comprehensive search for dark matter annihilation in dwarf galaxies," *Phys. Rev. D*, vol. 91, p. 083 535, 8 Apr. 2015. DOI: 10.1103/PhysRevD.91.083535. URL: https://link.aps.org/doi/10.1103/PhysRevD.91.083535.

[101] M. Ackermann *et al.*, "Dark matter constraints from observations of 25 milky way satellite galaxies with the fermi large area telescope," *Phys. Rev. D*, vol. 89, p. 042 001, 4 Feb. 2014. DOI: 10.1103/PhysRevD.89.042001. URL: https://link.aps.org/doi/10.1103/PhysRevD.89.042001.

[102] G. Bertone, *Particle Dark Matter : Observations, Models and Searches*. 2010.

[103] P. Colín, V. Avila-Reese, and O. Valenzuela, "Substructure and Halo Density Profiles in a Warm Dark Matter Cosmology," *ApJ*, vol. 542, no. 2, pp. 622–630, Oct. 2000. DOI: 10.1086/317057. arXiv: astro-ph/0004115 [astro-ph].

[104] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky, and O. Ruchayskiy, "The haloes of bright satellite galaxies in a warm dark matter universe," *MNRAS*, vol. 420, no. 3, pp. 2318–2324, Mar. 2012. DOI: 10.1111/j.1365-2966.2011.20200.x. arXiv: 1104.2929 [astro-ph.CO].

[105] S. G. Carlsten, J. E. Greene, R. L. Beaton, S. Danieli, and J. P. Greco, "The Exploration of Local VolumE Satellites (ELVES) Survey: A Nearly Volume-limited Sample of Nearby Dwarf Satellite Systems," *ApJ*, vol. 933, no. 1, 47, p. 47, Jul. 2022. DOI: 10.3847/1538-4357/ac6fd7. arXiv: 2203.00014 [astro-ph.GA].

[106] A. Helmi, "Streams, Substructures, and the Early History of the Milky Way," *ARA&A*, vol. 58, pp. 205–256, Aug. 2020. DOI: 10.1146/annurev-astro-032620-021917. arXiv: 2002.04340 [astro-ph.GA].

[107] DESI Collaboration, A. Aghamousa, J. Aguilar, *et al.*, "The DESI Experiment Part I: Science,Targeting, and Survey Design," *arXiv e-prints*, arXiv:1611.00036, arXiv:1611.00036, Oct. 2016. DOI: 10.48550/arXiv.1611.00036. arXiv: 1611.00036 [astro-ph.IM].

[108] J. Kwan, C. Sánchez, J. Clampitt, *et al.*, "Cosmology from large-scale galaxy clustering and galaxy-galaxy lensing with Dark Energy Survey Science Verification data," *MNRAS*, vol. 464, no. 4, pp. 4045–4062, Feb. 2017. DOI: 10.1093/mnras/stw2464. arXiv: 1604.07871 [astro-ph.CO].

[109] R. Mandelbaum, A. Slosar, T. Baldauf, U. Seljak, C. M. Hirata, R. Nakajima, R. Reyes, and R. E. Smith, "Cosmological parameter constraints from galaxy-galaxy lensing and galaxy clustering with the SDSS DR7," *MNRAS*, vol. 432, no. 2, pp. 1544–1575, Jun. 2013. DOI: 10.1093/mnras/stt572. arXiv: 1207.1120 [astro-ph.CO].

[110] J. P. Gardner, J. C. Mather, M. Clampin, *et al.*, "The James Webb Space Telescope," *Space Sci. Rev.*, vol. 123, no. 4, pp. 485–606, Apr. 2006. DOI: 10.1007/s11214-006-8315-7. arXiv: astro-ph/0606175 [astro-ph].

[111] S. L. Finkelstein *et al.*, "CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST," *ApJ*, vol. 946, no. 1, L13, p. L13, Mar. 2023. DOI: 10.3847/2041-8213/acade4. arXiv: 2211.05792 [astro-ph.GA].

[112] G. C. K. Leung *et al.*, "NGDEEP Epoch 1: The Faint End of the Luminosity Function at z 9-12 from Ultradeep JWST Imaging," *ApJ*, vol. 954, no. 2, L46, p. L46, Sep. 2023. DOI: 10.3847/2041-8213/acf365. arXiv: 2306.06244 [astro-ph.GA].

[113] Ž. Ivezić, S. M. Kahn, J. A. Tyson, *et al.*, "LSST: From Science Drivers to Reference Design and Anticipated Data Products," *ApJ*, vol. 873, 111, p. 111, Mar. 2019. DOI: 10.3847/1538-4357/ab042c. arXiv: 0805.2366.

[114] Q. Mao, A. A. Berlind, R. J. Scherrer, *et al.*, "A Cosmic Void Catalog of SDSS DR12 BOSS Galaxies," *ApJ*, vol. 835, no. 2, 161, p. 161, Feb. 2017. DOI: 10.3847/1538-4357/835/2/161. arXiv: 1602.02771 [astro-ph.CO].

[115] K. S. Dawson, D. J. Schlegel, C. P. Ahn, *et al.*, "The Baryon Oscillation Spectroscopic Survey of SDSS-III," *AJ*, vol. 145, no. 1, 10, p. 10, Jan. 2013. DOI: 10.1088/0004-6256/145/1/10. arXiv: 1208.0022 [astro-ph.CO].

[116] A. Del Popolo and M. Le Delliou, "Small scale problems of the ΛCDM model: a short review," *Galaxies*, vol. 5, no. 1, p. 17, 2017. DOI: 10.3390/galaxies5010017. arXiv: 1606.07790 [astro-ph.CO].

[117] J. F. Navarro, C. S. Frenk, and S. D. M. White, "A universal density profile from hierarchical clustering," *ApJ*, vol. 490, no. 2, pp. 493–508, Dec. 1997. DOI: 10.1086/304888. URL: https://doi.org/10.1086/304888.

[118] S.-H. Oh, D. A. Hunter, E. Brinks, *et al.*, "High-resolution Mass Models of Dwarf Galaxies from LITTLE THINGS," *AJ*, vol. 149, no. 6, 180, p. 180, Jun. 2015. DOI: 10.1088/0004-6256/149/6/180. arXiv: 1502.01281 [astro-ph.GA].

[119] J. J. Adams, J. D. Simon, M. H. Fabricius, *et al.*, "Dwarf Galaxy Dark Matter Density Profiles Inferred from Stellar and Gas Kinematics," *ApJ*, vol. 789, no. 1, 63, p. 63, Jul. 2014. DOI: 10.1088/0004-637X/789/1/63. arXiv: 1405.4854 [astro-ph.GA].

[120] K. Hayashi, M. Ibe, S. Kobayashi, Y. Nakayama, and S. Shirai, "Probing dark matter self-interaction with ultrafaint dwarf galaxies," *Phys. Rev. D*, vol. 103, no. 2, 023017, p. 023 017, Jan. 2021. DOI: 10.1103/PhysRevD.103.023017. arXiv: 2008.02529 [astro-ph.CO].

[121] L. V. Sales, A. Wetzel, and A. Fattahi, "Baryonic solutions and challenges for cosmological models of dwarf galaxies," *Nature Astronomy*, vol. 6, pp. 897–910, Jun. 2022. DOI: 10.1038/s41550-022-01689-w. arXiv: 2206.05295 [astro-ph.GA].

[122] K. A. Oman, J. F. Navarro, A. Fattahi, *et al.*, "The unexpected diversity of dwarf galaxy rotation curves," *MNRAS*, vol. 452, no. 4, pp. 3650–3665, Oct. 2015. DOI: 10.1093/mnras/stv1504. arXiv: 1504.01437 [astro-ph.GA].

[123] P. Creasey, O. Sameie, L. V. Sales, H.-B. Yu, M. Vogelsberger, and J. Zavala, "Spreading out and staying sharp - creating diverse rotation curves via baryonic and self-interaction effects," *MNRAS*, vol. 468, no. 2, pp. 2283–2295, Jun. 2017. DOI: 10.1093/mnras/stx522. arXiv: 1612.03903 [astro-ph.GA].

[124] J. Zavala, M. R. Lovell, M. Vogelsberger, and J. D. Burger, "Diverse dark matter density at sub-kiloparsec scales in Milky Way satellites: Implications for the nature of dark matter," *Phys. Rev. D*, vol. 100, no. 6, 063007, p. 063 007, Sep. 2019. DOI: 10.1103/PhysRevD.100.063007. arXiv: 1904.09998 [astro-ph.GA].

[125] K. Hayashi, M. Chiba, and T. Ishiyama, "Diversity of Dark Matter Density Profiles in the Galactic Dwarf Spheroidal Satellites," *ApJ*, vol. 904, no. 1, 45, p. 45, Nov. 2020. DOI: 10.3847/1538-4357/abbe0a. arXiv: 2007.13780 [astro-ph.GA].

[126] W. E. Kunkel and S. Demers, "The Magellanic Plane," in *The Galaxy and the Local Group*, vol. 182, Jan. 1976, p. 241.

[127] P. Kroupa, C. Theis, and C. M. Boily, "The great disk of Milky-Way satellites and cosmological sub-structures," *A&A*, vol. 431, pp. 517–521, Feb. 2005. DOI: 10.1051/0004-6361:20041122. arXiv: astro-ph/0410421 [astro-ph].

[128] M. S. Pawlowski, J. Pflamm-Altenburg, and P. Kroupa, "The VPOS: a vast polar structure of satellite galaxies, globular clusters and streams around the Milky Way," *MNRAS*, vol. 423, no. 2, pp. 1109–1126, Jun. 2012. DOI: 10.1111/j.1365-2966.2012.20937.x. arXiv: 1204.5176 [astro-ph.GA].

[129] A. R. Conn, G. F. Lewis, R. A. Ibata, *et al.*, "The Three-dimensional Structure of the M31 Satellite System; Strong Evidence for an Inhomogeneous Distribution of Satellites," *ApJ*, vol. 766, no. 2, 120, p. 120, Apr. 2013. DOI: 10.1088/0004-637X/766/2/120. arXiv: 1301.7131 [astro-ph.CO].

[130] R. A. Ibata, G. F. Lewis, A. R. Conn, *et al.*, "A vast, thin plane of corotating dwarf galaxies orbiting the Andromeda galaxy," *Nature*, vol. 493, no. 7430, pp. 62–65, Jan. 2013. DOI: 10.1038/nature11717. arXiv: 1301.0446 [astro-ph.CO].

[131] B. Mutlu-Pakdil, D. J. Sand, D. Crnojević, A. Drlica-Wagner, N. Caldwell, P. Guhathakurta, A. C. Seth, J. D. Simon, J. Strader, and E. Toloba, "Resolved Dwarf Galaxy Searches within 5 Mpc with the Vera Rubin Observatory and Subaru Hyper Suprime-Cam," *ApJ*, vol. 918, no. 2, 88, p. 88, Sep. 2021. DOI: 10.3847/1538-4357/ac0db8. arXiv: 2105.01658 [astro-ph.GA].

[132] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, "Too big to fail? The puzzling darkness of massive Milky Way subhaloes," *MNRAS*, vol. 415, no. 1, pp. L40–L44, Jul. 2011. DOI: 10.1111/j.1745-3933.2011.01074.x. arXiv: 1103.0007 [astro-ph.CO].

[133] E. Papastergis, R. Giovanelli, M. P. Haynes, and F. Shankar, "Is there a "too big to fail" problem in the field?" *A&A*, vol. 574, A113, A113, Feb. 2015. DOI: 10.1051/0004-6361/201424909. arXiv: 1407.4665 [astro-ph.GA].

[134] X. Kang, L. Wang, and Y. Luo, "The gap of stellar mass in galaxy groups: another perspective of the too-big-to-fail problem in the Milky Way," *MNRAS*, vol. 460, no. 2, pp. 2152–2156, Aug. 2016. DOI: 10.1093/mnras/stw1166. arXiv: 1605.03964 [astro-ph.GA].

[135] E. Papastergis and F. Shankar, "An assessment of the "too big to fail" problem for field dwarf galaxies in view of baryonic feedback effects," *A&A*, vol. 591, A58, A58, Jun. 2016. DOI: 10.1051/0004-6361/201527854. arXiv: 1511.08741 [astro-ph.GA].

[136] A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, "Where Are the Missing Galactic Satellites?" *ApJ*, vol. 522, no. 1, pp. 82–92, Sep. 1999. DOI: 10.1086/307643. arXiv: astro-ph/9901240 [astro-ph].

[137] E. J. Tollerud, J. S. Bullock, L. E. Strigari, and B. Willman, "Hundreds of Milky Way Satellites? Luminosity Bias in the Satellite Luminosity Function," *ApJ*, vol. 688, no. 1, pp. 277–289, Nov. 2008. DOI: 10.1086/592102. arXiv: 0806.4381 [astro-ph].

[138] A. Drlica-Wagner, K. Bechtol, S. Mau, *et al.*, "Milky Way Satellite Census. I. The Observational Selection Function for Milky Way Satellites in DES Y3 and Pan-STARRS DR1," *ApJ*, vol. 893, no. 1, 47, p. 47, Apr. 2020. DOI: 10.3847/1538-4357/ab7eb9. arXiv: 1912.03302 [astro-ph.GA].

[139] A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert, "Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-mass Galaxy," *ApJ*, vol. 827, no. 2, L23, p. L23, Aug. 2016. DOI: 10.3847/2041-8205/827/2/L23. arXiv: 1602.05957 [astro-ph.GA].

[140] A. Fitts, M. Boylan-Kolchin, O. D. Elbert, *et al.*, "fire in the field: simulating the threshold of galaxy formation," *MNRAS*, vol. 471, no. 3, pp. 3547–3562, Nov. 2017. DOI: 10.1093/mnras/stx1757. arXiv: 1611.02281 [astro-ph.GA].

[141] S. Garrison-Kimmel, A. Wetzel, J. S. Bullock, *et al.*, "Not so lumpy after all: modelling the depletion of dark matter subhaloes by Milky Way-like galaxies," *MNRAS*, vol. 471, no. 2, pp. 1709–1727, Oct. 2017. DOI: 10.1093/mnras/stx1710. arXiv: 1701.03792 [astro-ph.GA].

[142] S. Y. Kim, A. H. G. Peter, and J. R. Hargis, "Missing Satellites Problem: Completeness Corrections to the Number of Satellite Galaxies in the Milky Way are Consistent with Cold Dark Matter Predictions," *Phys. Rev. Lett.*, vol. 121, no. 21, 211302, p. 211 302, Nov. 2018. DOI: 10.1103/PhysRevLett.121.211302. arXiv: 1711.06267 [astro-ph.CO].

[143] C. M. Simpson, R. J. J. Grand, F. A. Gómez, F. Marinacci, R. Pakmor, V. Springel, D. J. R. Campbell, and C. S. Frenk, "Quenching and ram pressure stripping of simulated Milky Way satellite galaxies," *MNRAS*, vol. 478, no. 1, pp. 548–567, Jul. 2018. DOI: 10.1093/mnras/sty774. arXiv: 1705.03018 [astro-ph.GA].

[144] S. Garrison-Kimmel, P. F. Hopkins, A. Wetzel, *et al.*, "The Local Group on FIRE: dwarf galaxy populations across a suite of hydrodynamic simulations," *MNRAS*, vol. 487, no. 1, pp. 1380–1399, Jul. 2019. DOI: 10.1093/mnras/stz1317. arXiv: 1806.04143 [astro-ph.GA].

[145] T. Buck, A. V. Macciò, A. A. Dutton, A. Obreja, and J. Frings, "NIHAO XV: the environmental impact of the host galaxy on galactic satellite and field dwarf galaxies," *MNRAS*, vol. 483, no. 1, pp. 1314–1341, Feb. 2019. DOI: 10.1093/mnras/sty2913. arXiv: 1804.04667 [astro-ph.GA].

[146] L. Mayer, C. Mastropietro, J. Wadsley, J. Stadel, and B. Moore, "Simultaneous ram pressure and tidal stripping; how dwarf spheroidals lost their gas," *MNRAS*, vol. 369, no. 3, pp. 1021–1038, Jul. 2006. DOI: 10.1111/j.1365-2966.2006.10403.x. arXiv: astro-ph/0504277 [astro-ph].

[147] M. S. Pawlowski, B. Famaey, D. Merritt, and P. Kroupa, "On the Persistence of Two Small-scale Problems in ΛCDM," *ApJ*, vol. 815, no. 1, 19, p. 19, Dec. 2015. DOI: 10.1088/0004-637X/815/1/19. arXiv: 1510.08060 [astro-ph.GA].

[148] I. M. E. Santos-Santos, J. F. Navarro, A. Robertson, A. Benítez-Llambay, K. A. Oman, M. R. Lovell, C. S. Frenk, A. D. Ludlow, A. Fattahi, and A. Ritz, "Baryonic clues to the puzzling diversity of dwarf galaxy rotation curves," *MNRAS*, vol. 495, no. 1, pp. 58–77, Jun. 2020. DOI: 10.1093/mnras/staa1072. arXiv: 1911.09116 [astro-ph.GA].

[149] M. Vogelsberger, F. Marinacci, P. Torrey, and E. Puchwein, "Cosmological simulations of galaxy formation," *Nature Reviews Physics*, vol. 2, no. 1, pp. 42–66, Jan. 2020. DOI: 10.1038/s42254-019-0127-2. arXiv: 1909.07976 [astro-ph.GA].

[150] V. Springel, N. Yoshida, and S. D. White, "Gadget: A code for collisionless and gasdynamical cosmological simulations," *New Astronomy*, vol. 6, no. 2, pp. 79–117, 2001, ISSN: 1384-1076. DOI: https://doi.org/10.1016/S1384-1076(01)00042-2. URL: https://www.sciencedirect.com/science/article/pii/S1384107601000422.

[151] V. Springel, "The cosmological simulation code gadget-2," *Monthly Notices of the Royal Astronomical Society*, vol. 364, no. 4, pp. 1105–1134, Dec. 2005, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09655.x. eprint: https://academic.oup.com/mnras/article-pdf/364/4/1105/18657201/364-4-1105.pdf. URL: https://doi.org/10.1111/j.1365-2966.2005.09655.x.

[152] A. A. Klypin, S. Trujillo-Gomez, and J. Primack, "Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation," *ApJ*, vol. 740, no. 2, 102, p. 102, Oct. 2011. DOI: 10.1088/0004-637X/740/2/102. arXiv: 1002.3660 [astro-ph.CO].

[153] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß, "MultiDark simulations: the story of dark matter halo concentrations and density profiles," *Monthly Notices of the Royal Astronomical Society*, vol. 457, no. 4, pp. 4340–4359, Feb. 2016, ISSN: 0035-8711. DOI: 10.1093/mnras/stw248. eprint: https://academic.oup.com/mnras/article-pdf/457/4/4340/18515365/stw248.pdf. URL: https://doi.org/10.1093/mnras/stw248.

[154] L. Y. A. Yung, R. S. Somerville, T. Nguyen, C. Modi, and J. P. Gardner, *The gureft simulations – dark matter halo demographics and assembly histories at ultrahigh redshift*, 2023.

[155] P. F. Hopkins, A. Wetzel, D. Kereš, *et al.*, "FIRE-2 simulations: physics versus numerics in galaxy formation," *MNRAS*, vol. 480, no. 1, pp. 800–863, Oct. 2018. DOI: 10.1093/mnras/sty1690. arXiv: 1702.06148 [astro-ph.GA].

[156] R. J. J. Grand, F. A. Gómez, F. Marinacci, R. Pakmor, V. Springel, D. J. R. Campbell, C. S. Frenk, A. Jenkins, and S. D. M. White, "The Auriga Project: the properties and formation mechanisms of disc galaxies across cosmic time," *MNRAS*, vol. 467, no. 1, pp. 179–207, May 2017. DOI: 10.1093/mnras/stx071. arXiv: 1610.01159 [astro-ph.GA].

[157] L. Wang, A. A. Dutton, G. S. Stinson, A. V. Macciò, C. Penzo, X. Kang, B. W. Keller, and J. Wadsley, "NIHAO project - I. Reproducing the inefficiency of galaxy formation across cosmic time with a large sample of cosmological hydrodynamical simulations," *MNRAS*, vol. 454, no. 1, pp. 83–94, Nov. 2015. DOI: 10.1093/mnras/stv1937. arXiv: 1503.04818 [astro-ph.GA].

[158] T. Sawala, C. S. Frenk, A. Fattahi, *et al.*, "The APOSTLE simulations: solutions to the Local Group's cosmic puzzles," *MNRAS*, vol. 457, no. 2, pp. 1931–1943, Apr. 2016. DOI: 10.1093/mnras/stw145. arXiv: 1511.01098 [astro-ph.GA].

[159] D. Nelson, V. Springel, A. Pillepich, *et al.*, "The IllustrisTNG simulations: public data release," *Computational Astrophysics and Cosmology*, vol. 6, no. 1, 2, p. 2, May 2019. DOI: 10.1186/s40668-019-0028-x. arXiv: 1812.05609 [astro-ph.GA].

[160] A. Wetzel, C. C. Hayward, R. E. Sanderson, *et al.*, "Public Data Release of the FIRE-2 Cosmological Zoom-in Simulations of Galaxy Formation," *ApJS*, vol. 265, no. 2, 44, p. 44, Apr. 2023. DOI: 10.3847/1538-4365/acb99a. arXiv: 2202.06969 [astro-ph.GA].

[161] V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann, "Populating a cluster of galaxies - I. Results at z=0," *MNRAS*, vol. 328, no. 3, pp. 726–750, Dec. 2001. DOI: 10.1046/j.1365-8711.2001.04912.x. arXiv: astro-ph/0012055 [astro-ph].

[162] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, "The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores," *The Astrophysical Journal*, vol. 762, no. 2, p. 109, Dec. 2012. DOI: 10.1088/0004-637X/762/2/109. URL: https://dx.doi.org/10.1088/0004-637X/762/2/109.

[163] S. R. Knollmann and A. Knebe, "AHF: Amiga's Halo Finder," *ApJS*, vol. 182, no. 2, pp. 608–624, Jun. 2009. DOI: 10.1088/0067-0049/182/2/608. arXiv: 0904.3662 [astro-ph.CO].

[164] A. Knebe, S. R. Knollmann, S. I. Muldrew, *et al.*, "Haloes gone MAD: The Halo-Finder Comparison Project," *MNRAS*, vol. 415, no. 3, pp. 2293–2318, Aug. 2011. DOI: 10.1111/j.1365-2966.2011.18858.x. arXiv: 1104.0949 [astro-ph.CO].

[165] J. Onions, A. Knebe, F. R. Pearce, *et al.*, "Subhaloes going Notts: the subhalo-finder comparison project," *MNRAS*, vol. 423, no. 2, pp. 1200–1214, Jun. 2012. DOI: 10.1111/j.1365-2966.2012.20947.x. arXiv: 1203.3695 [astro-ph.CO].

[166] P. Mansfield, E. Darragh-Ford, Y. Wang, E. O. Nadler, and R. H. Wechsler, "Symfind: Addressing the Fragility of Subhalo Finders and Revealing the Durability of Subhalos," *arXiv e-prints*, arXiv:2308.10926, arXiv:2308.10926, Aug. 2023. DOI: 10.48550/arXiv.2308.10926. arXiv: 2308.10926 [astro-ph.CO].

[167] C. Scannapieco, M. Wadepuhl, O. H. Parry, *et al.*, "The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation," *MNRAS*, vol. 423, no. 2, pp. 1726–1749, Jun. 2012. DOI: 10.1111/j.1365-2966.2012.20993.x. arXiv: 1112.0315 [astro-ph.GA].

[168] J.-h. Kim, T. Abel, O. Agertz, *et al.*, "The AGORA High-resolution Galaxy Simulations Comparison Project," *ApJS*, vol. 210, no. 1, 14, p. 14, Jan. 2014. DOI: 10.1088/0067-0049/210/1/14. arXiv: 1308.2669 [astro-ph.GA].

[169] E. Zhang, L. V. Sales, F. Marinacci, P. Torrey, M. Vogelsberger, V. Springel, H. Li, R. Pakmor, and T. A. Gutcke, "Bursty Star Formation in Dwarfs is Sensitive to Numerical Choices in Supernova Feedback Models," *arXiv e-prints*, arXiv:2406.10338, arXiv:2406.10338, Jun. 2024. DOI: 10.48550/arXiv.2406.10338. arXiv: 2406.10338 [astro-ph.GA].

[170] S. Huang, N. Katz, R. Davé, *et al.*, "The robustness of cosmological hydrodynamic simulation predictions to changes in numerics and cooling physics," *MNRAS*, vol. 484, no. 2, pp. 2021–2046, Apr. 2019. DOI: 10.1093/mnras/stz057. arXiv: 1810.12946 [astro-ph.GA].

[171] F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, *et al.*, "The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations," *ApJ*, vol. 915, no. 1, 71, p. 71, Jul. 2021. DOI: 10.3847/1538-4357/abf7ba. arXiv: 2010.00619 [astro-ph.CO].

[172] J. C. Rose, P. Torrey, F. Villaescusa-Navarro, *et al.*, "Introducing the DREAMS Project: DaRk mattEr and Astrophysics with Machine learning and Simulations," *arXiv e-prints*, arXiv:2405.00766, arXiv:2405.00766, May 2024. DOI: 10.48550/arXiv.2405.00766. arXiv: 2405.00766 [astro-ph.GA].

[173] R. E. Sanderson, A. Wetzel, S. Loebman, S. Sharma, P. F. Hopkins, S. Garrison-Kimmel, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert, "Synthetic Gaia Surveys from the FIRE Cosmological Simulations of Milky Way-mass Galaxies," *ApJS*,

vol. 246, no. 1, 6, p. 6, Jan. 2020. DOI: 10.3847/1538-4365/ab5b9d. arXiv: 1806.10564 [astro-ph.GA].

[174] T. Nguyen, X. Ou, N. Panithanpaisal, N. Shipp, L. Necib, R. Sanderson, and A. Wetzel, "Synthetic Gaia DR3 Surveys from the FIRE Cosmological Simulations of Milky Way-mass Galaxies," *ApJ*, vol. 966, no. 1, 108, p. 108, May 2024. DOI: 10.3847/1538-4357/ad35ba. arXiv: 2306.16475 [astro-ph.GA].

[175] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander, "The MICE grand challenge lightcone simulation - I. Dark matter clustering," *MNRAS*, vol. 448, no. 4, pp. 2987–3000, Apr. 2015. DOI: 10.1093/mnras/stv138. arXiv: 1312.1707 [astro-ph.CO].

[176] R. Overzier, G. Lemson, R. E. Angulo, E. Bertin, J. Blaizot, B. M. B. Henriques, G. .-. Marleau, and S. D. M. White, "The Millennium Run Observatory: first light," *MNRAS*, vol. 428, no. 1, pp. 778–803, Jan. 2013. DOI: 10.1093/mnras/sts076. arXiv: 1206.6923 [astro-ph.CO].

[177] M. Manera, R. Scoccimarro, W. J. Percival, *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues," *MNRAS*, vol. 428, no. 2, pp. 1036–1054, Jan. 2013. DOI: 10.1093/mnras/sts084. arXiv: 1203.6609 [astro-ph.CO].

[178] Y.-S. Ting, C. Conroy, H.-W. Rix, and P. Cargile, "The Payne: Self-consistent ab initio Fitting of Stellar Spectra," *ApJ*, vol. 879, no. 2, 69, p. 69, Jul. 2019. DOI: 10.3847/1538-4357/ab2331. arXiv: 1804.01530 [astro-ph.SR].

[179] R. Ahumada, C. Allende Prieto, A. Almeida, *et al.*, "The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra," *ApJS*, vol. 249, no. 1, 3, p. 3, Jul. 2020. DOI: 10.3847/1538-4365/ab929e. arXiv: 1912.02905 [astro-ph.GA].

[180] T. M. C. Abbott, M. Adamów, M. Aguena, *et al.*, "The Dark Energy Survey Data Release 2," *ApJS*, vol. 255, no. 2, 20, p. 20, Aug. 2021. DOI: 10.3847/1538-4365/ac00b3. arXiv: 2101.05765 [astro-ph.IM].

[181] P. Sartoretti, O. Marchal, C. Babusiaux, *et al.*, "Gaia Data Release 3: G_RVS photometry from the RVS spectra," *arXiv e-prints*, arXiv:2206.05725, arXiv:2206.05725, Jun. 2022. arXiv: 2206.05725 [astro-ph.GA].

[182] D. Katz, P. Sartoretti, A. Guerrier, *et al.*, "Gaia Data Release 3 Properties and validation of the radial velocities," *arXiv e-prints*, arXiv:2206.05902, arXiv:2206.05902, Jun. 2022. arXiv: 2206.05902 [astro-ph.GA].

[183] M. Fouesneau, Y. Frémat, R. Andrae, *et al.*, "Gaia Data Release 3: Apsis II – Stellar Parameters," *arXiv e-prints*, arXiv:2206.05992, arXiv:2206.05992, Jun. 2022. DOI: 10.48550/arXiv.2206.05992. arXiv: 2206.05992 [astro-ph.SR].

[184] M. Riello, F. De Angeli, D. W. Evans, *et al.*, "Gaia Early Data Release 3. Photometric content and validation," *A&A*, vol. 649, A3, A3, May 2021. DOI: 10.1051/0004-6361/202039587. arXiv: 2012.01916 [astro-ph.IM].

[185] T. Isobe, E. D. Feigelson, M. G. Akritas, and G. J. Babu, "Linear Regression in Astronomy. I.," *ApJ*, vol. 364, p. 104, Nov. 1990. DOI: 10.1086/169390.

[186] G. Ashton *et al.*, "BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy," *ApJS*, vol. 241, no. 2, 27, p. 27, Apr. 2019. DOI: 10.3847/1538-4365/ab06fc. arXiv: 1811.02042 [astro-ph.IM].

[187] Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, *et al.*, "First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole," *ApJ*, vol. 875, no. 1, L4, p. L4, Apr. 2019. DOI: 10.3847/2041-8213/ab0e85. arXiv: 1906.11241 [astro-ph.GA].

[188] Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, *et al.*, "First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole," *ApJ*, vol. 875, no. 1, L6, p. L6, Apr. 2019. DOI: 10.3847/2041-8213/ab1141. arXiv: 1906.11243 [astro-ph.GA].

[189] L. E. Strigari, S. M. Koushiappas, J. S. Bullock, M. Kaplinghat, J. D. Simon, M. Geha, and B. Willman, "The Most Dark-Matter-dominated Galaxies: Predicted Gamma-Ray Signals from the Faintest Milky Way Dwarfs," *ApJ*, vol. 678, no. 2, pp. 614–620, May 2008. DOI: 10.1086/529488. arXiv: 0709.1510 [astro-ph].

[190] J. Rüstig, M. Guardiani, J. Roth, P. Frank, and T. Enßlin, "Introducing LensCharm. A charming Bayesian strong lensing reconstruction framework," *A&A*, vol. 682, A146, A146, Feb. 2024. DOI: 10.1051/0004-6361/202348256.

[191] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, "X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue," *Astron. Astrophys.*, vol. 564, A125, 2014. DOI: 10.1051/0004-6361/201322971. arXiv: 1402.0004 [astro-ph.HE].

[192] S. Mondal and L. Mastrolorenzo, "Machine Learning in High Energy Physics: A review of heavy-flavor jet tagging at the LHC," *arXiv e-prints*, arXiv:2404.01071, arXiv:2404.01071, Apr. 2024. DOI: 10.48550/arXiv.2404.01071. arXiv: 2404.01071 [hep-ex].

[193] A. Butter, T. Plehn, S. Schumann, *et al.*, "Machine learning and LHC event generation," *SciPost Physics*, vol. 14, no. 4, 079, p. 079, Apr. 2023. DOI: 10.21468/SciPostPhys.14.4.079. arXiv: 2203.07460 [hep-ph].

[194] A. Gagliano, G. Contardo, D. Foreman-Mackey, A. I. Malz, and P. D. Aleo, "First Impressions: Early-time Classification of Supernovae Using Host-galaxy Information and Shallow Learning," *ApJ*, vol. 954, no. 1, 6, p. 6, Sep. 2023. DOI: 10.3847/1538-4357/ace326. arXiv: 2305.08894 [astro-ph.IM].

[195] E. Marx, W. Benoit, A. Gunny, *et al.*, "A machine-learning pipeline for real-time detection of gravitational waves from compact binary coalescences," *arXiv e-prints*, arXiv:2403.18661, arXiv:2403.18661, Mar. 2024. DOI: 10.48550/arXiv.2403.18661. arXiv: 2403.18661 [gr-qc].

[196] S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *MNRAS*, vol. 450, no. 2, pp. 1441–1459, Jun. 2015. DOI: 10.1093/mnras/stv632. arXiv: 1503.07077 [astro-ph.IM].

[197] M. Zevin and Gravity Spy, "Gravity Spy - Integrating LIGO detector characterization, citizen science, and machine learning," in *American Astronomical Society Meeting Abstracts #228*, ser. American Astronomical Society Meeting Abstracts, vol. 228, Jun. 2016, 109.02, p. 109.02.

[198] A. Gagliano and V. A. Villar, "A Physics-Informed Variational Autoencoder for Rapid Galaxy Inference and Anomaly Detection," *arXiv e-prints*, arXiv:2312.16687, arXiv:2312.16687, Dec. 2023. DOI: 10.48550/arXiv.2312.16687. arXiv: 2312.16687 [astro-ph.IM].

[199] L. Parker, F. Lanusse, S. Golkar, *et al.*, "AstroCLIP: A Cross-Modal Foundation Model for Galaxies," *arXiv e-prints*, arXiv:2310.03024, arXiv:2310.03024, Oct. 2023. DOI: 10.48550/arXiv.2310.03024. arXiv: 2310.03024 [astro-ph.IM].

[200] A. D'Isanto and K. L. Polsterer, "Photometric redshift estimation via deep learning. Generalized and pre-classification-less, image based, fully probabilistic redshifts," *A&A*, vol. 609, A111, A111, Jan. 2018. DOI: 10.1051/0004-6361/201731326. arXiv: 1706.02467 [astro-ph.IM].

[201] J. C. Rose, P. Torrey, F. Villaescusa-Navarro, M. Vogelsberger, S. O'Neil, M. V. Medvedev, R. Low, R. Adhikari, and D. Anglés-Alcázar, "Inferring warm dark matter masses with deep learning," *MNRAS*, vol. 527, no. 1, pp. 739–755, Jan. 2024. DOI: 10.1093/mnras/stad3260. arXiv: 2304.14432 [astro-ph.CO].

[202] P. Villanueva-Domingo and F. Villaescusa-Navarro, "Learning Cosmology and Clustering with Cosmic Graphs," *ApJ*, vol. 937, no. 2, 115, p. 115, Oct. 2022. DOI: 10.3847/1538-4357/ac8930. arXiv: 2204.13713 [astro-ph.CO].

[203] N. S. M. de Santi, H. Shao, F. Villaescusa-Navarro, *et al.*, "Robust Field-level Likelihood-free Inference with Galaxies," *ApJ*, vol. 952, no. 1, 69, p. 69, Jul. 2023. DOI: 10.3847/1538-4357/acd1e2. arXiv: 2302.14101 [astro-ph.CO].

[204] T. L. Makinen, T. Charnock, P. Lemos, N. Porqueres, A. Heavens, and B. D. Wandelt, "The Cosmic Graph: Optimal Information Extraction from Large-Scale Structure using Catalogues," *arXiv e-prints*, arXiv:2207.05202, arXiv:2207.05202, Jul. 2022. arXiv: 2207.05202 [astro-ph.CO].

[205] S. Wagner-Carena, J. Aalbers, S. Birrer, E. O. Nadler, E. Darragh-Ford, P. J. Marshall, and R. H. Wechsler, "From Images to Dark Matter: End-to-end Inference of Substructure from Hundreds of Strong Gravitational Lenses," *ApJ*, vol. 942, no. 2, 75, p. 75, Jan. 2023. DOI: 10.3847/1538-4357/aca525. arXiv: 2203.00690 [astro-ph.CO].

[206] S. Wagner-Carena, J. Lee, J. Pennington, J. Aalbers, S. Birrer, and R. H. Wechsler, "A Strong Gravitational Lens Is Worth a Thousand Dark Matter Halos: Inference on Small-Scale Structure Using Sequential Methods," *arXiv e-prints*, arXiv:2404.14487, arXiv:2404.14487, Apr. 2024. DOI: 10.48550/arXiv.2404.14487. arXiv: 2404.14487 [astro-ph.CO].

[207] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, "Real-Time Gravitational Wave Science with Neural Posterior Estimation," *Phys. Rev. Lett.*, vol. 127, no. 24, 241103, p. 241 103, Dec. 2021. DOI: 10.1103/PhysRevLett.127.241103. arXiv: 2106.12594 [gr-qc].

[208] R. Raikman, E. A. Moreno, E. Govorkova, *et al.*, "GWAK: gravitational-wave anomalous knowledge with recurrent autoencoders," *Machine Learning: Science and Technology*, vol. 5, no. 2, 025020, p. 025 020, Jun. 2024. DOI: 10.1088/2632-2153/ad3a31. arXiv: 2309.11537 [astro-ph.IM].

[209] K. Storey-Fisher, M. Huertas-Company, N. Ramachandra, F. Lanusse, A. Leauthaud, Y. Luo, and S. Huang, "Anomaly Detection in Astronomical Images with Generative Adversarial Networks," *arXiv e-prints*, arXiv:2012.08082, arXiv:2012.08082, Dec. 2020. DOI: 10.48550/arXiv.2012.08082. arXiv: 2012.08082 [astro-ph.GA].

[210] M. Bussov and J. Nättilä, "Segmentation of turbulent computational fluid dynamics simulations with unsupervised ensemble learning," *Signal Processing: Image Communication*, vol. 99, 116450, p. 116 450, Nov. 2021. DOI: 10.1016/j.image.2021.116450. arXiv: 2109.01381 [cs.CV].

[211] A. Akhmetzhanova, S. Mishra-Sharma, and C. Dvorkin, "Data compression and inference in cosmology with self-supervised machine learning," *MNRAS*, vol. 527, no. 3, pp. 7459–7481, Jan. 2024. DOI: 10.1093/mnras/stad3646. arXiv: 2308.09751 [astro-ph.CO].

[212] D. Schaurecker, Y. Li, J. Tinker, S. Ho, and A. Refregier, "Super-resolving Dark Matter Halos using Generative Deep Learning," *arXiv e-prints*, arXiv:2111.06393, arXiv:2111.06393, Nov. 2021. DOI: 10.48550/arXiv.2111.06393. arXiv: 2111.06393 [astro-ph.CO].

[213] D. Kodi Ramanah, T. Charnock, and G. Lavaux, "Painting halos from cosmic density fields of dark matter with physically motivated neural networks," *Phys. Rev. D*, vol. 100, no. 4, 043515, p. 043 515, Aug. 2019. DOI: 10.1103/PhysRevD.100.043515. arXiv: 1903.10524 [astro-ph.CO].

[214] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, "Noise reduction in gravitational-wave data via deep learning," *Physical Review Research*, vol. 2, no. 3, 033066, p. 033 066, Jul. 2020. DOI: 10.1103/PhysRevResearch.2.033066. arXiv: 2005.06534 [astro-ph.IM].

[215] M. Saleem, A. Gunny, C.-J. Chou, *et al.*, "Demonstration of Machine Learning-assisted real-time noise regression in gravitational wave detectors," *arXiv e-prints*, arXiv:2306.11366, arXiv:2306.11366, Jun. 2023. DOI: 10.48550/arXiv.2306.11366. arXiv: 2306.11366 [gr-qc].

[216] K. Boone, "ParSNIP: Generative Models of Transient Light Curves with Physics-enabled Deep Learning," *AJ*, vol. 162, no. 6, 275, p. 275, Dec. 2021. DOI: 10.3847/1538-3881/ac2a2d. arXiv: 2109.13999 [astro-ph.IM].

[217] G. Papamakarios, E. Nalisnick, D. Jimenez Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing Flows for Probabilistic Modeling and Inference," *arXiv e-prints*, arXiv:1912.02762, arXiv:1912.02762, Dec. 2019. DOI: 10.48550/arXiv.1912.02762. arXiv: 1912.02762 [stat.ML].

[218] C. C. Lovell, S. Hassan, F. Villaescusa-Navarro, *et al.*, "A Hierarchy of Normalizing Flows for Modelling the Galaxy-Halo Relationship," in *Machine Learning for Astrophysics*, Jul. 2023, 21, p. 21. DOI: 10.48550/arXiv.2307.06967. arXiv: 2307.06967 [astro-ph.GA].

[219] C. Cuesta-Lazaro and S. Mishra-Sharma, "A point cloud approach to generative modeling for galaxy surveys at the field level," *arXiv e-prints*, arXiv:2311.17141, arXiv:2311.17141, Nov. 2023. DOI: 10.48550/arXiv.2311.17141. arXiv: 2311.17141 [astro-ph.CO].

[220] A. Ćiprijanović, D. Kafkes, S. Jenkins, K. Downey, G. N. Perdue, S. Madireddy, T. Johnston, and B. Nord, "Domain adaptation techniques for improved cross-domain study of galaxy mergers," *arXiv e-prints*, arXiv:2011.03591, arXiv:2011.03591, Nov. 2020. DOI: 10.48550/arXiv.2011.03591. arXiv: 2011.03591 [astro-ph.IM].

[221] F. Ott, D. Rügamer, L. Heublein, B. Bischl, and C. Mutschler, "Domain Adaptation for Time-Series Classification to Mitigate Covariate Shift," *arXiv e-prints*, arXiv:2204.03342, arXiv:2204.03342, Apr. 2022. DOI: 10.48550/arXiv.2204.03342. arXiv: 2204.03342 [cs.LG].

[222] X. Liu, C. Yoo, F. Xing, H. Oh, G. El Fakhri, J.-W. Kang, and J. Woo, "Deep Unsupervised Domain Adaptation: A Review of Recent Advances and Perspectives," *arXiv e-prints*, arXiv:2208.07422, arXiv:2208.07422, Aug. 2022. DOI: 10.48550/arXiv.2208.07422. arXiv: 2208.07422 [cs.CV].

[223] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv e-prints*, arXiv:1705.07874, arXiv:1705.07874, May 2017. DOI: 10.48550/arXiv.1705.07874. arXiv: 1705.07874 [cs.AI].

[224] L. Scorzato, "Reliability and Interpretability in Science and Deep Learning," *arXiv e-prints*, arXiv:2401.07359, arXiv:2401.07359, Jan. 2024. DOI: 10.48550/arXiv.2401.07359. arXiv: 2401.07359 [cs.AI].

[225] F. Fan, J. Xiong, M. Li, and G. Wang, "On Interpretability of Artificial Neural Networks: A Survey," *arXiv e-prints*, arXiv:2001.02522, arXiv:2001.02522, Jan. 2020. DOI: 10.48550/arXiv.2001.02522. arXiv: 2001.02522 [cs.LG].

[226] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[227] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv e-prints*, arXiv:1502.03167, arXiv:1502.03167, Feb. 2015. DOI: 10.48550/arXiv.1502.03167. arXiv: 1502.03167 [cs.LG].

[228] T. Chen, "On the Importance of Noise Scheduling for Diffusion Models," *arXiv e-prints*, arXiv:2301.10972, arXiv:2301.10972, Jan. 2023. DOI: 10.48550/arXiv.2301.10972. arXiv: 2301.10972 [cs.CV].

[229] P. Nakkiran, G. Kaplun, Y. Bansal, T. Yang, B. Barak, and I. Sutskever, "Deep Double Descent: Where Bigger Models and More Data Hurt," *arXiv e-prints*, arXiv:1912.02292, arXiv:1912.02292, Dec. 2019. DOI: 10.48550/arXiv.1912.02292. arXiv: 1912.02292 [cs.LG].

[230] Y. Gu, X. Zheng, and T. Aste, "Unraveling the Enigma of Double Descent: An In-depth Analysis through the Lens of Learned Feature Space," *arXiv e-prints*, arXiv:2310.13572, arXiv:2310.13572, Oct. 2023. DOI: 10.48550/arXiv.2310.13572. arXiv: 2310.13572 [cs.LG].

[231] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proceedings of the National Academy of Science*, vol. 117, no. 48, pp. 30 055–30 062, Dec. 2020. DOI: 10.1073/pnas.1912789117. arXiv: 1911.01429 [stat.ML].

[232] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, F. R. Bach and D. M. Blei, Eds., ser. JMLR Workshop and Conference Proceedings, vol. 37, 2015, pp. 1530–1538. URL: http://proceedings.mlr.press/v37/rezende15.html.

[233] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," *arXiv e-prints*, arXiv:2006.11239, arXiv:2006.11239, Jun. 2020. DOI: 10.48550/arXiv.2006.11239. arXiv: 2006.11239 [cs.LG].

[234] C. Luo, "Understanding Diffusion Models: A Unified Perspective," *arXiv e-prints*, arXiv:2208.11970, arXiv:2208.11970, Aug. 2022. DOI: 10.48550/arXiv.2208.11970. arXiv: 2208.11970 [cs.LG].

[235] D. P. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational Diffusion Models," *arXiv e-prints*, arXiv:2107.00630, arXiv:2107.00630, Jul. 2021. DOI: 10.48550/arXiv.2107.00630. arXiv: 2107.00630 [cs.LG].

[236] D. P. Kingma and R. Gao, "Understanding Diffusion Objectives as the ELBO with Simple Data Augmentation," *arXiv e-prints*, arXiv:2303.00848, arXiv:2303.00848, Mar. 2023. DOI: 10.48550/arXiv.2303.00848. arXiv: 2303.00848 [cs.LG].

[237] R. Durall, A. Chatzimichailidis, P. Labus, and J. Keuper, "Combating Mode Collapse in GAN training: An Empirical Analysis using Hessian Eigenvalues," *arXiv e-prints*, arXiv:2012.09673, arXiv:2012.09673, Dec. 2020. DOI: 10.48550/arXiv.2012.09673. arXiv: 2012.09673 [cs.LG].

[238] H. Liu, B. Li, H. Wu, H. Liang, Y. Huang, Y. Li, B. Ghanem, and Y. Zheng, "Combating Mode Collapse in GANs via Manifold Entropy Estimation," *arXiv e-prints*, arXiv:2208.12055, arXiv:2208.12055, Aug. 2022. DOI: 10.48550/arXiv.2208.12055. arXiv: 2208.12055 [cs.CV].

[239] N. Mudur and D. P. Finkbeiner, "Can denoising diffusion probabilistic models generate realistic astrophysical fields?" *arXiv e-prints*, arXiv:2211.12444, arXiv:2211.12444, Nov. 2022. DOI: 10.48550/arXiv.2211.12444. arXiv: 2211.12444 [astro-ph.CO].

[240] X. Zhao, Y.-S. Ting, K. Diao, and Y. Mao, "Can diffusion model conditionally generate astrophysical images?" *MNRAS*, vol. 526, no. 2, pp. 1699–1712, Dec. 2023. DOI: 10.1093/mnras/stad2778. arXiv: 2307.09568 [astro-ph.IM].

[241] M. Drozdova, V. Kinakh, O. Bait, O. Taran, E. Lastufka, M. Dessauges-Zavadsky, T. Holotyak, D. Schaerer, and S. Voloshynovskiy, "Radio-astronomical image reconstruction with a conditional denoising diffusion model," *A&A*, vol. 683, A105, A105, Mar. 2024. DOI: 10.1051/0004-6361/202347948. arXiv: 2402.10204 [astro-ph.IM].

[242] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *arXiv e-prints*, arXiv:1710.10903, arXiv:1710.10903, Oct. 2017. arXiv: 1710.10903 [stat.ML].

[243] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. arXiv: 1903.02428 [cs.LG].

[244] L.-H. Lim and B. J. Nelson, "What is an equivariant neural network?" *arXiv e-prints*, arXiv:2205.07362, arXiv:2205.07362, May 2022. DOI: 10.48550/arXiv.2205.07362. arXiv: 2205.07362 [cs.LG].

[245] X. Zhang, J. Yoon, M. Bansal, and H. Yao, "Multimodal Representation Learning by Alternating Unimodal Adaptation," *arXiv e-prints*, arXiv:2311.10707, arXiv:2311.10707, Nov. 2023. DOI: 10.48550/arXiv.2311.10707. arXiv: 2311.10707 [cs.LG].

[246] Z. Lu, "A Theory of Multimodal Learning," *arXiv e-prints*, arXiv:2309.12458, arXiv:2309.12458, Sep. 2023. DOI: 10.48550/arXiv.2309.12458. arXiv: 2309.12458 [cs.LG].

[247] Y. Zong, O. Mac Aodha, and T. Hospedales, "Self-Supervised Multimodal Learning: A Survey," *arXiv e-prints*, arXiv:2304.01008, arXiv:2304.01008, Mar. 2023. DOI: 10.48550/arXiv.2304.01008. arXiv: 2304.01008 [cs.LG].

[248] Y. Zhang, K. Gong, K. Zhang, H. Li, Y. Qiao, W. Ouyang, and X. Yue, "Meta-Transformer: A Unified Framework for Multimodal Learning," *arXiv e-prints*, arXiv:2307.10802, arXiv:2307.10802, Jul. 2023. DOI: 10.48550/arXiv.2307.10802. arXiv: 2307.10802 [cs.CV].

[249] I. V. Slijepcevic, A. M. M. Scaife, M. Walmsley, M. Bowles, O. I. Wong, S. S. Shabala, and S. V. White, "Radio galaxy zoo: towards building the first multipurpose foundation model for radio astronomy with self-supervised learning," *RAS Techniques and Instruments*, vol. 3, no. 1, pp. 19–32, Jan. 2024. DOI: 10.1093/rasti/rzad055. arXiv: 2305.16127 [astro-ph.IM].

[250] T. Nguyen, C. Modi, L. Y. A. Yung, and R. S. Somerville, "FLORAH: A generative model for halo assembly histories," *arXiv e-prints*, arXiv:2308.05145, arXiv:2308.05145, Aug. 2023. DOI: 10.48550/arXiv.2308.05145. arXiv: 2308.05145 [astro-ph.GA].

[251] A. Lazar, J. S. Bullock, M. Boylan-Kolchin, *et al.*, "A dark matter profile to model diverse feedback-induced core sizes of ΛCDM haloes," *MNRAS*, vol. 497, no. 2, pp. 2393–2417, Sep. 2020. DOI: 10.1093/mnras/staa2101. arXiv: 2004.10817 [astro-ph.GA].

[252] A. McDaniel, M. Ajello, C. M. Karwin, M. Di Mauro, A. Drlica-Wagner, and M. A. Sánchez-Conde, "Legacy analysis of dark matter annihilation from the Milky Way dwarf spheroidal galaxies with 14 years of Fermi -LAT data," *Phys. Rev. D*, vol. 109, no. 6, 063024, p. 063 024, Mar. 2024. DOI: 10.1103/PhysRevD.109.063024. arXiv: 2311.04982 [astro-ph.HE].

[253] P. F. Hopkins, A. Wetzel, D. Kereš, *et al.*, "FIRE-2 simulations: physics versus numerics in galaxy formation," *MNRAS*, vol. 480, no. 1, pp. 800–863, Oct. 2018. DOI: 10.1093/mnras/sty1690. arXiv: 1702.06148 [astro-ph.GA].

[254] L. J. Chang and L. Necib, "Dark matter density profiles in dwarf galaxies: linking Jeans modelling systematics and observation," *Monthly Notices of the Royal Astronomical Society*, vol. 507, no. 4, pp. 4715–4733, Aug. 2021, ISSN: 0035-8711. DOI: 10.1093/mnras/stab2440. eprint: https://academic.oup.com/mnras/article-pdf/507/4/4715/40384382/stab2440.pdf. URL: https://doi.org/10.1093/mnras/stab2440.

[255] J. Neyman and E. L. Scott, "A Theory of the Spatial Distribution of Galaxies.," *ApJ*, vol. 116, p. 144, Jul. 1952. DOI: 10.1086/145599.

[256] A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh, and C. G. Lacey, "The nature of galaxy bias and clustering," *MNRAS*, vol. 311, no. 4, pp. 793–808, Feb. 2000. DOI: 10.1046/j.1365-8711.2000.03101.x. arXiv: astro-ph/9903343 [astro-ph].

[257] J. A. Peacock and R. E. Smith, "Halo occupation numbers and galaxy bias," *MNRAS*, vol. 318, no. 4, pp. 1144–1156, Nov. 2000. DOI: 10.1046/j.1365-8711.2000.03779.x. arXiv: astro-ph/0005010 [astro-ph].

[258] A. P. Hearin, A. R. Zentner, F. C. van den Bosch, D. Campbell, and E. Tollerud, "Introducing decorated HODs: modelling assembly bias in the galaxy-halo connection," *MNRAS*, vol. 460, no. 3, pp. 2552–2570, Aug. 2016. DOI: 10.1093/mnras/stw840. arXiv: 1512.03050 [astro-ph.CO].

[259] X. Yang, H. J. Mo, and F. C. van den Bosch, "Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies," *MNRAS*, vol. 339, no. 4, pp. 1057–1080, Mar. 2003. DOI: 10.1046/j.1365-8711.2003.06254.x. arXiv: astro-ph/0207019 [astro-ph].

[260] C. Conroy and R. H. Wechsler, "Connecting Galaxies, Halos, and Star Formation Rates Across Cosmic Time," *ApJ*, vol. 696, no. 1, pp. 620–635, May 2009. DOI: 10.1088/0004-637X/696/1/620. arXiv: 0805.3346 [astro-ph].

[261] D. J. Croton, "A simple model to link the properties of quasars to the properties of dark matter haloes out to high redshift," *MNRAS*, vol. 394, no. 2, pp. 1109–1119, Apr. 2009. DOI: 10.1111/j.1365-2966.2009.14429.x. arXiv: 0901.4104 [astro-ph.CO].

[262] A. Smith, C. Grove, S. Cole, *et al.*, "Generating mock galaxy catalogues for flux-limited samples like the DESI Bright Galaxy Survey," *MNRAS*, vol. 532, no. 1, pp. 903–919, Jul. 2024. DOI: 10.1093/mnras/stae1503. arXiv: 2312.08792 [astro-ph.CO].

[263] N. Hamaus, M. Aubert, A. Pisani, *et al.*, "Euclid: Forecasts from redshift-space distortions and the Alcock-Paczynski test with cosmic voids," *A&A*, vol. 658, A20, A20, Feb. 2022. DOI: 10.1051/0004-6361/202142073. arXiv: 2108.10347 [astro-ph.CO].

[264] S. L. Finkelstein, G. C. K. Leung, M. B. Bagley, *et al.*, "The Complete CEERS Early Universe Galaxy Sample: A Surprisingly Slow Evolution of the Space Density of Bright Galaxies at z $\sim$ 8.5–14.5," *ApJ*, vol. 969, no. 1, L2, p. L2, Jul. 2024. DOI: 10.3847/2041-8213/ad4495. arXiv: 2311.04279 [astro-ph.GA].

[265] G. Kauffmann, J. M. Colberg, A. Diaferio, and S. D. M. White, "Clustering of galaxies in a hierarchical universe — I. Methods and results at z = 0," *Monthly Notices of the Royal Astronomical Society*, vol. 303, no. 1, pp. 188–206, Feb. 1999, ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02202.x. eprint: https://academic.oup.com/mnras/

article-pdf/303/1/188/3940810/303-1-188.pdf. URL: https://doi.org/10.1046/j.1365-8711.1999.02202.x.

[266] R. S. Somerville, G. Lemson, Y. Sigad, A. Dekel, G. Kauffmann, and S. D. M. White, "Non-linear stochastic galaxy biasing in cosmological simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 320, no. 3, pp. 289–306, Jan. 2001, ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.03894.x. eprint: https://academic.oup.com/mnras/article-pdf/320/3/289/3793114/320-3-289.pdf. URL: https://doi.org/10.1046/j.1365-8711.2001.03894.x.

[267] A. J. Benson, R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole, "What Shapes the Luminosity Function of Galaxies?" *ApJ*, vol. 599, no. 1, pp. 38–49, Dec. 2003. DOI: 10.1086/379160. arXiv: astro-ph/0302450 [astro-ph].

[268] R. S. Somerville, C. Olsen, L. Y. A. Yung, *et al.*, "Mock light-cones and theory friendly catalogues for the CANDELS survey," *MNRAS*, vol. 502, no. 4, pp. 4858–4876, Apr. 2021. DOI: 10.1093/mnras/stab231. arXiv: 2102.00108 [astro-ph.GA].

[269] B. Hadzhiyska, S. Liu, R. S. Somerville, A. Gabrielpillai, S. Bose, D. Eisenstein, and L. Hernquist, "Galaxy assembly bias and large-scale distribution: a comparison between IllustrisTNG and a semi-analytic model," *MNRAS*, vol. 508, no. 1, pp. 698–718, Nov. 2021. DOI: 10.1093/mnras/stab2564. arXiv: 2108.00006 [astro-ph.CO].

[270] M. Ayromlou, G. Kauffmann, R. M. Yates, D. Nelson, and S. D. M. White, "Galaxy formation with L-GALAXIES: modelling the environmental dependency of galaxy evolution and comparing with observations," *MNRAS*, vol. 505, no. 1, pp. 492–514, Jul. 2021. DOI: 10.1093/mnras/stab1245. arXiv: 2011.05336 [astro-ph.GA].

[271] A. Gabrielpillai, R. S. Somerville, S. Genel, V. Rodriguez-Gomez, V. Pandya, L. Y. A. Yung, and L. Hernquist, "Galaxy formation in the Santa Cruz semi-analytic model compared with IllustrisTNG – I. Galaxy scaling relations, dispersions, and residuals at z = 0," *Monthly Notices of the Royal Astronomical Society*, vol. 517, no. 4, pp. 6091–6111, Aug. 2022, ISSN: 0035-8711. DOI: 10.1093/mnras/stac2297. eprint: https://academic.oup.com/mnras/article-pdf/517/4/6091/47012908/stac2297.pdf. URL: https://doi.org/10.1093/mnras/stac2297.

[272] L. Y. A. Yung, R. S. Somerville, H. C. Ferguson, S. L. Finkelstein, J. P. Gardner, R. Davé, M. B. Bagley, G. Popping, and P. Behroozi, "Semi-analytic forecasts for JWST - VI. Simulated light-cones and galaxy clustering predictions," *MNRAS*, vol. 515, no. 4, pp. 5416–5436, Oct. 2022. DOI: 10.1093/mnras/stac2139. arXiv: 2206.13521 [astro-ph.GA].

[273] L. Y. A. Yung, R. S. Somerville, S. L. Finkelstein, *et al.*, "Semi-analytic forecasts for Roman - the beginning of a new era of deep-wide galaxy surveys," *MNRAS*, vol. 519, no. 1, pp. 1578–1600, Feb. 2023. DOI: 10.1093/mnras/stac3595. arXiv: 2210.04902 [astro-ph.GA].

[274] C. d. P. Lagos, E. Bayet, C. M. Baugh, C. G. Lacey, T. A. Bell, N. Fanidakis, and J. E. Geach, "Predictions for the CO emission of galaxies from a coupled simulation of galaxy formation and photon-dominated regions," *MNRAS*, vol. 426, no. 3, pp. 2142–2165, Nov. 2012. DOI: 10.1111/j.1365-2966.2012.21905.x. arXiv: 1204.0795 [astro-ph.CO].

[275] C. d. P. Lagos, C. G. Lacey, and C. M. Baugh, "A dynamical model of supernova feedback: gas outflows from the interstellar medium," *MNRAS*, vol. 436, no. 2, pp. 1787–1817, Dec. 2013. DOI: 10.1093/mnras/stt1696. arXiv: 1303.6635 [astro-ph.CO].

[276] R. S. Somerville, G. Popping, and S. C. Trager, "Star formation in semi-analytic galaxy formation models with multiphase gas," *Monthly Notices of the Royal Astronomical Society*, vol. 453, no. 4, pp. 4337–4367, Sep. 2015, ISSN: 0035-8711. DOI: 10.1093/mnras/stv1877. eprint: https://academic.oup.com/mnras/article-pdf/453/4/4337/8034633/stv1877.pdf. URL: https://doi.org/10.1093/mnras/stv1877.

[277] G. Popping, R. S. Somerville, and M. Galametz, "The dust content of galaxies from z = 0 to z = 9," *MNRAS*, vol. 471, no. 3, pp. 3152–3185, Nov. 2017. DOI: 10.1093/mnras/stx1545. arXiv: 1609.08622 [astro-ph.GA].

[278] L. Y. A. Yung, R. S. Somerville, G. Popping, S. L. Finkelstein, H. C. Ferguson, and R. Davé, "Semi-analytic forecasts for JWST - II. Physical properties and scaling relations for galaxies at z = 4-10," *MNRAS*, vol. 490, no. 2, pp. 2855–2879, Dec. 2019. DOI: 10.1093/mnras/stz2755. arXiv: 1901.05964 [astro-ph.GA].

[279] L. Y. A. Yung, R. S. Somerville, G. Popping, and S. L. Finkelstein, "Semi-analytic forecasts for JWST - III. Intrinsic production efficiency of Lyman-continuum radiation," *MNRAS*, vol. 494, no. 1, pp. 1002–1017, May 2020. DOI: 10.1093/mnras/staa714. arXiv: 1910.11345 [astro-ph.GA].

[280] B. M. B. Henriques, R. M. Yates, J. Fu, Q. Guo, G. Kauffmann, C. Srisawat, P. A. Thomas, and S. D. M. White, "L-GALAXIES 2020: Spatially resolved cold gas phases, star formation, and chemical enrichment in galactic discs," *MNRAS*, vol. 491, no. 4, pp. 5795–5814, Feb. 2020. DOI: 10.1093/mnras/stz3233. arXiv: 2003.05944 [astro-ph.GA].

[281] R. M. Yates, B. M. B. Henriques, J. Fu, G. Kauffmann, P. A. Thomas, Q. Guo, S. D. M. White, and P. Schady, "L-GALAXIES 2020: The evolution of radial metallicity profiles and global metallicities in disc galaxies," *MNRAS*, vol. 503, no. 3, pp. 4474–4495, May 2021. DOI: 10.1093/mnras/stab741. arXiv: 2011.04670 [astro-ph.GA].

[282] R. S. Somerville, P. F. Hopkins, T. J. Cox, B. E. Robertson, and L. Hernquist, "A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei," *Monthly Notices of the Royal Astronomical Society*, vol. 391, no. 2, pp. 481–506, Nov. 2008, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2008.13805.x. eprint: https://academic.oup.com/mnras/article-pdf/391/2/481/5764146/mnras0391-0481.pdf. URL: https://doi.org/10.1111/j.1365-2966.2008.13805.x.

[283] N. Fanidakis, C. M. Baugh, A. J. Benson, R. G. Bower, S. Cole, C. Done, and C. S. Frenk, "Grand unification of AGN activity in the ΛCDM cosmology," *MNRAS*, vol. 410, no. 1, pp. 53–74, Jan. 2011. DOI: 10.1111/j.1365-2966.2010.17427.x. arXiv: 0911.1128 [astro-ph.CO].

[284] L. Y. A. Yung, R. S. Somerville, S. L. Finkelstein, M. Hirschmann, R. Davé, G. Popping, J. P. Gardner, and A. Venkatesan, "Semi-analytic forecasts for JWST - V. AGN luminosity functions and helium reionization at z = 2-7," *MNRAS*, vol. 508, no. 2, pp. 2706–2729, Dec. 2021. DOI: 10.1093/mnras/stab2761. arXiv: 2109.13241 [astro-ph.GA].

[285] A. Hutter, P. Dayal, G. Yepes, S. Gottlöber, L. Legrand, and G. Ucci, "Astraeus I: the interplay between galaxy formation and reionization," *MNRAS*, vol. 503, no. 3, pp. 3698–3723, May 2021. DOI: 10.1093/mnras/stab602. arXiv: 2004.08401 [astro-ph.GA].

[286] G. Ucci, P. Dayal, A. Hutter, G. Yepes, S. Gottlöber, L. Legrand, L. Pentericci, M. Castellano, and T. R. Choudhury, "Astraeus - II. Quantifying the impact of cosmic variance during the Epoch of Reionization," *MNRAS*, vol. 506, no. 1, pp. 202–214, Sep. 2021. DOI: 10.1093/mnras/stab1229. arXiv: 2004.11096 [astro-ph.GA].

[287] A. Hutter, P. Dayal, L. Legrand, S. Gottlöber, and G. Yepes, "Astraeus - III. The environment and physical properties of reionization sources," *MNRAS*, vol. 506, no. 1, pp. 215–228, Sep. 2021. DOI: 10.1093/mnras/stab877. arXiv: 2008.13215 [astro-ph.GA].

[288] S. D. M. White and M. J. Rees, "Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering.," *MNRAS*, vol. 183, pp. 341–358, May 1978. DOI: 10.1093/mnras/183.3.341.

[289] J. F. Navarro, V. R. Eke, and C. S. Frenk, "The cores of dwarf galaxy haloes," *MNRAS*, vol. 283, no. 3, pp. L72–L78, Dec. 1996. DOI: 10.1093/mnras/283.3.L72. arXiv: astro-ph/9610187 [astro-ph].

[290] K. Spekkens, R. Giovanelli, and M. P. Haynes, "The Cusp/Core Problem in Galactic Halos: Long-Slit Spectra for a Large Dwarf Galaxy Sample," *AJ*, vol. 129, no. 5, pp. 2119–2137, May 2005. DOI: 10.1086/429592. arXiv: astro-ph/0502166 [astro-ph].

[291] J. H. Jeans, "On the theory of star-streaming and the structure of the universe," *MNRAS*, vol. 76, pp. 70–84, Dec. 1915. DOI: 10.1093/mnras/76.2.70.

[292] A. Genina, J. I. Read, C. S. Frenk, S. Cole, A. Benítez-Llambay, A. D. Ludlow, J. F. Navarro, K. A. Oman, and A. Robertson, "To β or not to β: can higher order Jeans analysis break the mass–anisotropy degeneracy in simulated dwarfs?" *Monthly Notices of the Royal Astronomical Society*, vol. 498, no. 1, pp. 144–163, Aug. 2020, ISSN: 0035-8711. DOI: 10.1093/mnras/staa2352. eprint: https://academic.oup.com/mnras/article-pdf/498/1/144/33703538/staa2352.pdf. URL: https://doi.org/10.1093/mnras/staa2352.

[293] J. I. Read, G. A. Mamon, E. Vasiliev, L. L. Watkins, M. G. Walker, J. Peñarrubia, M. Wilkinson, W. Dehnen, and P. Das, "Breaking beta: a comparison of mass modelling methods for spherical systems," *MNRAS*, vol. 501, no. 1, pp. 978–993, Feb. 2021. DOI: 10.1093/mnras/staa3663. arXiv: 2011.09493 [astro-ph.GA].

[294] J. Read, M. Walker, and P. Steger, "The case for a cold dark matter cusp in Draco," *MNRAS*, vol. 481, no. 1, pp. 860–877, 2018. DOI: 10.1093/mnras/sty2286. arXiv: 1805.06934 [astro-ph.GA].

[295] M. G. Walker and J. Peñarrubia, "A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies," *ApJ*, vol. 742, 20, p. 20, Nov. 2011. DOI: 10.1088/0004-637X/742/1/20. arXiv: 1108.2404.

[296] N. C. Amorisco and N. W. Evans, "Dark matter cores and cusps: the case of multiple stellar populations in dwarf spheroidals," *MNRAS*, vol. 419, no. 1, pp. 184–196, Jan. 2012. DOI: 10.1111/j.1365-2966.2011.19684.x. arXiv: 1106.1062 [astro-ph.CO].

[297] L. Zhu, G. van de Ven, L. L. Watkins, and L. Posti, "A discrete chemo-dynamical model of the dwarf spheroidal galaxy Sculptor: mass profile, velocity anisotropy and internal rotation," *MNRAS*, vol. 463, no. 1, pp. 1117–1135, Nov. 2016. DOI: 10.1093/mnras/stw2081. arXiv: 1608.08239 [astro-ph.GA].

[298] L. E. Strigari, J. S. Bullock, and M. Kaplinghat, "Determining the Nature of Dark Matter with Astrometry," *ApJ Lett.*, vol. 657, pp. L1–L4, 2007. DOI: 10.1086/512976. arXiv: astro-ph/0701581.

[299] L. E. Strigari, C. S. Frenk, and S. D. White, "Dynamical constraints on the dark matter distribution of the Sculptor dwarf spheroidal from stellar proper motions," *ApJ*, vol. 860, no. 1, p. 56, 2018. DOI: 10.3847/1538-4357/aac2d3. arXiv: 1801.07343 [astro-ph.GA].

[300] G. A. Mamon, A. Biviano, and G. Boué, "MAMPOSSt: Modelling Anisotropy and Mass Profiles of Observed Spherical Systems - I. Gaussian 3D velocities," *MNRAS*, vol. 429, no. 4, pp. 3079–3098, Mar. 2013. DOI: 10.1093/mnras/sts565. arXiv: 1212.1455 [astro-ph.CO].

[301] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis* (Chapman & Hall/CRC Texts in Statistical Science). Chapman & Hall/CRC, 2003, ISBN: 9781584883883. URL: https://books.google.com/books?id=BkGCmAEACAAJ.

[302] A. F. M. Smith and A. E. Gelfand, "Bayesian statistics without tears: A sampling–resampling perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, 1992. DOI: 10.1080/00031305.1992.10475856. eprint: https://doi.org/10.1080/00031305.1992.10475856. URL: https://doi.org/10.1080/00031305.1992.10475856.

[303] D. B. Rubin, "Using the sir algorithm to simulate posterior distributions," *Bayesian statistics*, vol. 3, pp. 395–402, 1988.

[304] H. C. Plummer, "On the problem of distribution in globular star clusters," *MNRAS*, vol. 71, pp. 460–470, Mar. 1911. DOI: 10.1093/mnras/71.5.460.

[305] L. P. Osipkov, "Spherical systems of gravitating bodies with an ellipsoidal velocity distribution," *Pisma v Astronomicheskii Zhurnal*, vol. 5, pp. 77–80, Feb. 1979.

[306] D. Merritt, "Spherical stellar systems with spheroidal velocity distributions," *AJ*, vol. 90, pp. 1027–1037, Jun. 1985. DOI: 10.1086/113810.

[307] J. D. Simon and M. Geha, "The Kinematics of the Ultra-faint Milky Way Satellites: Solving the Missing Satellite Problem," *ApJ*, vol. 670, no. 1, pp. 313–331, Nov. 2007. DOI: 10.1086/521816. arXiv: 0706.0516 [astro-ph].

[308] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," *arXiv e-prints*, arXiv:1606.09375, arXiv:1606.09375, Jun. 2016. arXiv: 1606.09375 [cs.LG].

[309] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *Journal of Machine Learning Research*, 2019. arXiv: 1912.02762 [cs.LG].

[310] M. Germain, K. Gregor, I. Murray, and H. Larochelle, "MADE: masked autoencoder for distribution estimation," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, F. R. Bach and D. M. Blei, Eds., ser. JMLR Workshop and Conference Proceedings, vol. 37, 2015, pp. 881–889. URL: http://proceedings.mlr.press/v37/germain15.html.

[311] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference," *Proc. Nat. Acad. Sci.*, vol. 117, no. 48, pp. 30 055–30 062, 2020. DOI: 10.1073/pnas.1912789117. arXiv: 1911.01429 [stat.ML].

[312] K. Cranmer and G. Louppe, "Unifying generative models and exact likelihood- free inference with conditional bijections," *J. Brief Ideas*, Dec. 2016. DOI: 10.5281/zenodo.198541. URL: https://doi.org/10.5281/zenodo.198541.

[313] G. Papamakarios and I. Murray, "Fast ϵ-free inference of simulation models with bayesian conditional density estimation," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16, Barcelona, Spain: Curran Associates Inc., 2016, pp. 1036–1044, ISBN: 9781510838819. arXiv: 1605.06376 [stat.ML]. URL: https://proceedings.neurips.cc/paper/2016/hash/6aca97005c68f1206823815f66102Abstract.html.

[314] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[315] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. URL: https://openreview.net/forum?id=Bkg6RiCqY7.

[316] J. Skilling, "Nested Sampling," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, R. Fischer, R. Preuss, and U. V. Toussaint, Eds., ser. American Institute of Physics Conference Series, vol. 735, Nov. 2004, pp. 395–405. DOI: 10.1063/1.1835238.

[317] J. Skilling, "Nested sampling for general bayesian computation," *Bayesian Anal.*, vol. 1, no. 4, pp. 833–859, Dec. 2006. DOI: 10.1214/06-BA127. URL: http://dx.doi.org/10.1214/06-BA127.

[318] J. S. Speagle, "DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences," *MNRAS*, vol. 493, no. 3, pp. 3132–3158, Apr. 2020. DOI: 10.1093/mnras/staa278. arXiv: 1904.02180 [astro-ph.IM].

[319] F. Perez and B. E. Granger, "IPython: A System for Interactive Scientific Computing," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 21–29, Jan. 2007. DOI: 10.1109/MCSE.2007.53.

[320] T. Kluyver *et al.*, "Jupyter notebooks - a publishing format for reproducible computational workflows," in *ELPUB*, 2016.

[321] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[322] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, *nflows: Normalizing flows in PyTorch*, version v0.14, Nov. 2020. DOI: 10.5281/zenodo.4296287. URL: https://doi.org/10.5281/zenodo.4296287.

[323] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, *et al.*, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[324] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[325] W. Falcon *et al.*, *Pytorchlightning/pytorch-lightning: 0.7.6 release*, version 0.7.6, May 2020. DOI: 10.5281/zenodo.3828935. URL: https://doi.org/10.5281/zenodo.3828935.

[326] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, 2020. DOI: https://doi.org/10.1038/s41592-019-0686-2.

[327] S. Cole, A. Aragon-Salamanca, C. S. Frenk, J. F. Navarro, and S. E. Zepf, "A recipe for galaxy formation.," *MNRAS*, vol. 271, pp. 781–806, Dec. 1994. DOI: 10.1093/mnras/271.4.781. arXiv: astro-ph/9402001 [astro-ph].

[328] S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk, "Hierarchical galaxy formation," *MNRAS*, vol. 319, no. 1, pp. 168–204, Nov. 2000. DOI: 10.1046/j.1365-8711.2000.03879.x. arXiv: astro-ph/0007281 [astro-ph].

[329] L. Y. A. Yung, R. Somerville, S. Finkelstein, G. Popping, and R. Davé, "Semi-analytic forecasts for JWST - I. UV luminosity functions at z = 4-10," *MNRAS*, vol. 483, no. 3, pp. 2983–3006, Mar. 2019. DOI: 10.1093/mnras/sty3241. arXiv: 1803.09761 [astro-ph.GA].

[330] E. J. Elliott, C. M. Baugh, and C. G. Lacey, "Efficient exploration and calibration of a semi-analytical model of galaxy formation with deep learning," *MNRAS*, vol. 506, no. 3, pp. 4011–4030, Sep. 2021. DOI: 10.1093/mnras/stab1837. arXiv: 2103.01072 [astro-ph.GA].

[331] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, "UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10," *MNRAS*, vol. 488, no. 3, pp. 3143–3194, Sep. 2019. DOI: 10.1093/mnras/stz1182. arXiv: 1806.07893 [astro-ph.GA].

[332] P. Behroozi, C. Conroy, R. H. Wechsler, *et al.*, "The Universe at z > 10: predictions for JWST from the UNIVERSEMACHINE DR1," *MNRAS*, vol. 499, no. 4, pp. 5702–5718, Dec. 2020. DOI: 10.1093/mnras/staa3164. arXiv: 2007.04988 [astro-ph.GA].

[333] B. P. Moster, T. Naab, and S. D. M. White, "EMERGE - an empirical model for the formation of galaxies since z ∼ 10," *MNRAS*, vol. 477, no. 2, pp. 1822–1852, Jun. 2018. DOI: 10.1093/mnras/sty655. arXiv: 1705.05373 [astro-ph.GA].

[334] R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel, "Concentrations and Assembly Histories of Dark Matter Halos," *arXiv e-prints*, astro-ph/0111069, astro–ph/0111069, Nov. 2001. DOI: 10.48550/arXiv.astro-ph/0111069. arXiv: astro-ph/0111069 [astro-ph].

[335] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, "Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations," *ApJ*, vol. 379, p. 440, Oct. 1991. DOI: 10.1086/170520.

[336] R. G. Bower, "The evolution of groups of galaxies in the Press-Schechter formalism," *MNRAS*, vol. 248, pp. 332–352, Jan. 1991. DOI: 10.1093/mnras/248.2.332.

[337] C. Lacey and S. Cole, "Merger rates in hierarchical models of galaxy formation," *MNRAS*, vol. 262, no. 3, pp. 627–649, Jun. 1993. DOI: 10.1093/mnras/262.3.627.

[338] R. S. Somerville and T. S. Kolatt, "How to plant a merger tree," *MNRAS*, vol. 305, no. 1, pp. 1–14, May 1999, ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02154.x. eprint: https://academic.oup.com/mnras/article-pdf/305/1/1/3080773/305-1-1.pdf. URL: https://doi.org/10.1046/j.1365-8711.1999.02154.x.

[339] A. R. Zentner, "The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth," *International Journal of Modern Physics D*, vol. 16, no. 5, pp. 763–815, Jan. 2007. DOI: 10.1142/S0218271807010511. arXiv: astro-ph/0611454 [astro-ph].

[340] E. Neistein and A. Dekel, "Constructing merger trees that mimic N-body simulations," *MNRAS*, vol. 383, no. 2, pp. 615–626, Jan. 2008. DOI: 10.1111/j.1365-2966.2007.12570.x. arXiv: 0708.1599 [astro-ph].

[341] H. Parkinson, S. Cole, and J. Helly, "Generating dark matter halo merger trees," *MNRAS*, vol. 383, no. 2, pp. 557–564, Jan. 2008. DOI: 10.1111/j.1365-2966.2007.12517.x. arXiv: 0708.1382 [astro-ph].

[342] C. A. Correa, J. S. B. Wyithe, J. Schaye, and A. R. Duffy, "The accretion history of dark matter haloes - I. The physical origin of the universal function," *MNRAS*, vol. 450, no. 2, pp. 1514–1520, Jun. 2015. DOI: 10.1093/mnras/stv689. arXiv: 1409.5228 [astro-ph.GA].

[343] C. A. Correa, J. S. B. Wyithe, J. Schaye, and A. R. Duffy, "The accretion history of dark matter haloes - III. A physical model for the concentration-mass relation," *MNRAS*, vol. 452, no. 2, pp. 1217–1232, Sep. 2015. DOI: 10.1093/mnras/stv1363. arXiv: 1502.00391 [astro-ph.CO].

[344] R. S. Somerville, G. Lemson, T. S. Kolatt, and A. Dekel, "Evaluating approximations for halo merging histories," *MNRAS*, vol. 316, no. 3, pp. 479–490, Aug. 2000. DOI: 10.1046/j.1365-8711.2000.03467.x. arXiv: astro-ph/9807277 [astro-ph].

[345] Y. Li, H. J. Mo, F. C. van den Bosch, and W. P. Lin, "On the assembly history of dark matter haloes," *MNRAS*, vol. 379, no. 2, pp. 689–701, Aug. 2007. DOI: 10.1111/j.1365-2966.2007.11942.x. arXiv: astro-ph/0510372 [astro-ph].

[346] A. J. Benson, A. Ludlow, and S. Cole, "Halo concentrations from extended Press-Schechter merger histories," *MNRAS*, vol. 485, no. 4, pp. 5010–5020, Jun. 2019. DOI: 10.1093/mnras/stz695. arXiv: 1812.06026 [astro-ph.GA].

[347] F. Jiang and F. C. van den Bosch, "Generating merger trees for dark matter haloes: a comparison of methods," *MNRAS*, vol. 440, no. 1, pp. 193–207, May 2014. DOI: 10.1093/mnras/stu280. arXiv: 1311.5225 [astro-ph.CO].

[348] S. Robles, J. S. Gómez, A. Ramírez Rivera, N. D. Padilla, and D. Dujovne, "A deep learning approach to halo merger tree construction," *MNRAS*, vol. 514, no. 3, pp. 3692–3708, Aug. 2022. DOI: 10.1093/mnras/stac1569. arXiv: 2205.15988 [astro-ph.GA].

[349] A. P. Hearin, J. Chaves-Montero, M. R. Becker, and A. Alarcon, "A Differentiable Model of the Assembly of Individual and Populations of Dark Matter Halos," *The Open Journal of Astrophysics*, vol. 4, no. 1, 7, p. 7, Jul. 2021. DOI: 10.21105/astro.2105.05859. arXiv: 2105.05859 [astro-ph.CO].

[350] L. Y. A. Yung, R. S. Somerville, S. L. Finkelstein, S. M. Wilkins, and J. P. Gardner, "Are the ultra-high-redshift galaxies at z $>$ 10 surprising in the context of standard galaxy formation models?" *arXiv e-prints*, arXiv:2304.04348, arXiv:2304.04348, Apr. 2023. DOI: 10.48550/arXiv.2304.04348. arXiv: 2304.04348 [astro-ph.GA].

[351] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores," *ApJ*, vol. 762, no. 2, 109, p. 109, Jan. 2013. DOI: 10.1088/0004-637X/762/2/109. arXiv: 1110.4372 [astro-ph.CO].

[352] P. S. Behroozi, R. H. Wechsler, H.-Y. Wu, M. T. Busha, A. A. Klypin, and J. R. Primack, "Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology," *ApJ*, vol. 763, no. 1, 18, p. 18, Jan. 2013. DOI: 10.1088/0004-637X/763/1/18. arXiv: 1110.4370 [astro-ph.CO].

[353] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, Eds., ser. Proceedings of Machine Learning Research, vol. 28, Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1310–1318. URL: https://proceedings.mlr.press/v28/pascanu13.html.

[354] A. H. Ribeiro, K. Tiels, L. A. Aguirre, and T. Schön, "Beyond exploding and vanishing gradients: Analysing rnn training using attractors and smoothness," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, S. Chiappa and R. Calandra, Eds., ser. Proceedings of Machine Learning Research, vol. 108, PMLR, 26–28 Aug 2020, pp. 2370–2380. URL: https://proceedings.mlr.press/v108/ribeiro20a.html.

[355] J. F. Navarro, C. S. Frenk, and S. D. M. White, "The Structure of Cold Dark Matter Halos," *ApJ*, vol. 462, p. 563, May 1996. DOI: 10.1086/177173. arXiv: astro-ph/9508025 [astro-ph].

[356] A. Nusser and R. K. Sheth, "Mass growth and density profiles of dark matter haloes in hierarchical clustering," *MNRAS*, vol. 303, no. 4, pp. 685–695, Mar. 1999. DOI: 10.1046/j.1365-8711.1999.02197.x. arXiv: astro-ph/9803281 [astro-ph].

[357] F. C. van den Bosch, "The universal mass accretion history of cold dark matter haloes," *MNRAS*, vol. 331, no. 1, pp. 98–110, Mar. 2002. DOI: 10.1046/j.1365-8711.2002.05171.x. arXiv: astro-ph/0105158 [astro-ph].

[358] R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov, and B. Allgood, "The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation," *ApJ*, vol. 652, no. 1, pp. 71–84, Nov. 2006. DOI: 10.1086/507120. arXiv: astro-ph/0512416 [astro-ph].

[359] C. A. Correa, J. S. B. Wyithe, J. Schaye, and A. R. Duffy, "The accretion history of dark matter haloes - II. The connections with the mass power spectrum and the density profile," *MNRAS*, vol. 450, no. 2, pp. 1521–1537, Jun. 2015. DOI: 10.1093/mnras/stv697. arXiv: 1501.04382 [astro-ph.CO].

[360] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is all you need*, 2023. arXiv: 1706.03762 [cs.CL].

[361] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, p. 6085, 2018.

[362] G. Papamakarios, T. Pavlakou, and I. Murray, "Masked autoregressive flow for density estimation," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 2335–2344, ISBN: 9781510860964. URL: https://papers.nips.cc/paper/2017/hash/6c1da886822c67822bcf3679d04369fa-Abstract.html.

[363] J. McBride, O. Fakhouri, and C.-P. Ma, "Mass accretion rates and histories of dark matter haloes," *MNRAS*, vol. 398, no. 4, pp. 1858–1868, Oct. 2009. DOI: 10.1111/j.1365-2966.2009.15329.x. arXiv: 0902.3659 [astro-ph.CO].

[364] D. Foreman-Mackey, "Corner.py: Scatterplot matrices in python," *The Journal of Open Source Software*, vol. 1, no. 2, p. 24, Jun. 2016. DOI: 10.21105/joss.00024. URL: https://doi.org/10.21105/joss.00024.

[365] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, Feb. 2020. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121 [cs.MS].

[366] B. D. Smith and M. Lang, "Ytree: A python package for analyzing merger trees," *Journal of Open Source Software*, vol. 4, no. 44, p. 1881, Dec. 2019. DOI: 10.21105/joss.01881. URL: https://doi.org/10.21105/joss.01881.

[367] R. H. Wechsler and J. L. Tinker, "The Connection Between Galaxies and Their Dark Matter Halos," *ARA&A*, vol. 56, pp. 435–487, Sep. 2018. DOI: 10.1146/annurev-astro-081817-051756. arXiv: 1804.03097 [astro-ph.GA].

[368] K. J. Kwon and C. Hahn, "Modeling the Kinematics of Central and Satellite Galaxies Using Normalizing Flows," *arXiv e-prints*, arXiv:2401.12318, arXiv:2401.12318, Jan. 2024. DOI: 10.48550/arXiv.2401.12318. arXiv: 2401.12318 [astro-ph.GA].

[369] D. Jimenez Rezende and S. Mohamed, "Variational Inference with Normalizing Flows," *arXiv e-prints*, arXiv:1505.05770, arXiv:1505.05770, May 2015. DOI: 10.48550/arXiv.1505.05770. arXiv: 1505.05770 [stat.ML].

[370] G. Papamakarios, T. Pavlakou, and I. Murray, "Masked autoregressive flow for density estimation," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 2335–2344, ISBN: 9781510860964. URL: https://papers.nips.cc/paper/2017/hash/6c1da886822c67822bcf3679d04369fa-Abstract.html.

[371] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations," *arXiv e-prints*, arXiv:2011.13456, arXiv:2011.13456, Nov. 2020. DOI: 10.48550/arXiv.2011.13456. arXiv: 2011.13456 [cs.LG].

[372] R. Weinberger, V. Springel, and R. Pakmor, "The AREPO Public Code Release," *ApJS*, vol. 248, no. 2, 32, p. 32, Jun. 2020. DOI: 10.3847/1538-4365/ab908c. arXiv: 1909.04667 [astro-ph.IM].

[373] R. Pakmor, A. Bauer, and V. Springel, "Magnetohydrodynamics on an unstructured moving grid," *MNRAS*, vol. 418, no. 2, pp. 1392–1401, Dec. 2011. DOI: 10.1111/j.1365-2966.2011.19591.x. arXiv: 1108.1792 [astro-ph.IM].

[374] R. Pakmor and V. Springel, "Simulations of magnetic fields in isolated disc galaxies," *MNRAS*, vol. 432, no. 1, pp. 176–193, Jun. 2013. DOI: 10.1093/mnras/stt428. arXiv: 1212.1452 [astro-ph.CO].

[375] O. Hahn and T. Abel, "Multi-scale initial conditions for cosmological simulations," *MNRAS*, vol. 415, no. 3, pp. 2101–2121, Aug. 2011. DOI: 10.1111/j.1365-2966.2011.18820.x. arXiv: 1103.6031 [astro-ph.CO].

[376] Planck Collaboration, P. A. R. Ade, N. Aghanim, *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *A&A*, vol. 594, A13, A13, Sep. 2016. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[377] D. Nelson, A. Pillepich, V. Springel, R. Pakmor, R. Weinberger, S. Genel, P. Torrey, M. Vogelsberger, F. Marinacci, and L. Hernquist, "First results from the TNG50 simulation: galactic outflows driven by supernovae and black hole feedback," *MNRAS*, vol. 490, no. 3, pp. 3234–3261, Dec. 2019. DOI: 10.1093/mnras/stz2306. arXiv: 1902.05554 [astro-ph.GA].

[378] A. Pillepich, D. Nelson, V. Springel, *et al.*, "First results from the TNG50 simulation: the evolution of stellar and gaseous discs across cosmic time," *MNRAS*, vol. 490, no. 3, pp. 3196–3233, Dec. 2019. DOI: 10.1093/mnras/stz2338. arXiv: 1902.05553 [astro-ph.GA].

[379] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist, "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe," *MNRAS*, vol. 444, no. 2, pp. 1518–1547, Oct. 2014. DOI: 10.1093/mnras/stu1536. arXiv: 1405.2921 [astro-ph.CO].

[380] P. Torrey, M. Vogelsberger, S. Genel, D. Sijacki, V. Springel, and L. Hernquist, "A model for cosmological simulations of galaxy formation physics: multi-epoch validation," *MNRAS*, vol. 438, no. 3, pp. 1985–2004, Mar. 2014. DOI: 10.1093/mnras/stt2295. arXiv: 1305.4931 [astro-ph.CO].

[381] I. M. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, vol. 7, no. 4, pp. 784–802, 1967.

[382] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*. 2008.

[383] H. Mo, F. C. van den Bosch, and S. White, *Galaxy Formation and Evolution*. 2010.

[384] S. Bose, D. J. Eisenstein, L. Hernquist, A. Pillepich, D. Nelson, F. Marinacci, V. Springel, and M. Vogelsberger, "Revealing the galaxy-halo connection in IllustrisTNG," *MNRAS*, vol. 490, no. 4, pp. 5693–5711, Dec. 2019. DOI: 10.1093/mnras/stz2546. arXiv: 1905.08799 [astro-ph.CO].

[385] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White, "The Aquarius Project: the subhaloes of galactic haloes," *MNRAS*, vol. 391, no. 4, pp. 1685–1711, Dec. 2008. DOI: 10.1111/j.1365-2966.2008.14066.x. arXiv: 0809.0898 [astro-ph].

[386] J. F. Navarro, C. S. Frenk, and S. D. M. White, "A Universal Density Profile from Hierarchical Clustering," *ApJ*, vol. 490, no. 2, pp. 493–508, Dec. 1997. DOI: 10.1086/304888. arXiv: astro-ph/9611107 [astro-ph].

[387] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, "Neural Spline Flows," *arXiv e-prints*, arXiv:1906.04032, arXiv:1906.04032, Jun. 2019. DOI: 10.48550/arXiv.1906.04032. arXiv: 1906.04032 [stat.ML].

[388] A. Vahdat, K. Kreis, and J. Kautz, "Score-based Generative Modeling in Latent Space," *arXiv e-prints*, arXiv:2106.05931, arXiv:2106.05931, Jun. 2021. DOI: 10.48550/arXiv.2106.05931. arXiv: 2106.05931 [stat.ML].

[389] Y. Song, C. Durkan, I. Murray, and S. Ermon, "Maximum Likelihood Training of Score-Based Diffusion Models," *arXiv e-prints*, arXiv:2101.09258, arXiv:2101.09258, Jan. 2021. DOI: 10.48550/arXiv.2101.09258. arXiv: 2101.09258 [stat.ML].

[390] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv e-prints*, arXiv:1312.6114, arXiv:1312.6114, Dec. 2013. DOI: 10.48550/arXiv.1312.6114. arXiv: 1312.6114 [stat.ML].

[391] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *arXiv e-prints*, arXiv:1406.2661, arXiv:1406.2661, Jun. 2014. DOI: 10.48550/arXiv.1406.2661. arXiv: 1406.2661 [stat.ML].

[392] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," *arXiv e-prints*, arXiv:1606.03498, arXiv:1606.03498, Jun. 2016. DOI: 10.48550/arXiv.1606.03498. arXiv: 1606.03498 [cs.LG].

[393] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[394] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv e-prints*, arXiv:1810.04805, arXiv:1810.04805, Oct. 2018. DOI: 10.48550/arXiv.1810.04805. arXiv: 1810.04805 [cs.CL].

[395] S. Y. Hwang, C. G. Sabiu, I. Park, and S. E. Hong, "The universe is worth 64^3 pixels: convolution neural network and vision transformers for cosmology," *J. Cosmology Astropart. Phys.*, vol. 2023, no. 11, 075, p. 075, Nov. 2023. DOI: 10.1088/1475-7516/2023/11/075. arXiv: 2304.08192 [astro-ph.CO].

[396] Y. He, J. Wu, W. Wang, B. Jiang, and Y. Zhang, "AstroYOLO: A hybrid CNN-Transformer deep-learning object-detection model for blue horizontal-branch stars," *PASJ*, vol. 75, no. 6, pp. 1311–1323, Dec. 2023. DOI: 10.1093/pasj/psad071.

[397] F. Lanusse, L. Parker, S. Golkar, *et al.*, "AstroCLIP: Cross-Modal Pre-Training for Astronomical Foundation Models," *arXiv e-prints*, arXiv:2310.03024, arXiv:2310.03024, Oct. 2023. DOI: 10.48550/arXiv.2310.03024. arXiv: 2310.03024 [astro-ph.IM].

[398] H. W. Leung and J. Bovy, "Towards an astronomical foundation model for stars with a transformer-based model," *MNRAS*, vol. 527, no. 1, pp. 1494–1520, Jan. 2024. DOI: 10.1093/mnras/stad3015. arXiv: 2308.10944 [astro-ph.IM].

[399] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)," *arXiv e-prints*, arXiv:1606.08415, arXiv:1606.08415, Jun. 2016. DOI: 10.48550/arXiv.1606.08415. arXiv: 1606.08415 [cs.LG].

[400] J. C. Rose, P. Torrey, F. Villaescusa-Navarro, *et al.*, "Introducing the DREAMS Project: DaRk mattEr and Astrophysics with Machine learning and Simulations," *arXiv e-prints*, arXiv:2405.00766, arXiv:2405.00766, May 2024. arXiv: 2405.00766 [astro-ph.GA].

[401] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. URL: https://openreview.net/forum?id=Bkg6RiCqY7.

[402] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[403] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," *arXiv e-prints*, arXiv:1608.03983, arXiv:1608.03983, Aug. 2016. DOI: 10.48550/arXiv.1608.03983. arXiv: 1608.03983 [cs.LG].

[404] M. R. Lovell, C. S. Frenk, V. R. Eke, A. Jenkins, L. Gao, and T. Theuns, "The properties of warm dark matter haloes," *MNRAS*, vol. 439, no. 1, pp. 300–317, Mar. 2014. DOI: 10.1093/mnras/stt2431. arXiv: 1308.1399 [astro-ph.CO].

[405] P. Jethwa, D. Erkal, and V. Belokurov, "The upper bound on the lowest mass halo," *MNRAS*, vol. 473, no. 2, pp. 2060–2083, Jan. 2018. DOI: 10.1093/mnras/stx2330. arXiv: 1612.07834 [astro-ph.GA].

[406] S. T. Brown, A. Fattahi, I. G. McCarthy, A. S. Font, K. A. Oman, and A. H. Riley, "ARTEMIS emulator: exploring the effect of cosmology and galaxy formation physics on Milky Way-mass haloes and their satellites," *arXiv e-prints*, arXiv:2403.11692, arXiv:2403.11692, Mar. 2024. DOI: 10.48550/arXiv.2403.11692. arXiv: 2403.11692 [astro-ph.GA].

[407] J. I. Read, G. Iorio, O. Agertz, and F. Fraternali, "The stellar mass-halo mass relation of isolated field dwarfs: a critical test of ΛCDM at the edge of galaxy formation," *MNRAS*, vol. 467, no. 2, pp. 2019–2038, May 2017. DOI: 10.1093/mnras/stx147. arXiv: 1607.03127 [astro-ph.GA].

[408] A. Di Cintio, C. B. Brook, A. V. Macciò, G. S. Stinson, A. Knebe, A. A. Dutton, and J. Wadsley, "The dependence of dark matter profiles on the stellar-to-halo mass ratio: a prediction for cusps versus cores," *MNRAS*, vol. 437, no. 1, pp. 415–423, Jan. 2014. DOI: 10.1093/mnras/stt1891. arXiv: 1306.0898 [astro-ph.CO].

[409] M. R. Lovell, W. Hellwing, A. Ludlow, J. Zavala, A. Robertson, A. Fattahi, C. S. Frenk, and J. Hardwick, "Local group star formation in warm and self-interacting dark matter cosmologies," *MNRAS*, vol. 498, no. 1, pp. 702–717, Oct. 2020. DOI: 10.1093/mnras/staa2525. arXiv: 2002.11129 [astro-ph.GA].

[410] C. C. Lovell, S. Hassan, F. Villaescusa-Navarro, *et al.*, "A Hierarchy of Normalizing Flows for Modelling the Galaxy-Halo Relationship," in *Machine Learning for Astrophysics*, Jul. 2023, 21, p. 21. DOI: 10.48550/arXiv.2307.06967. arXiv: 2307.06967 [astro-ph.GA].

[411] M. E. Lee, S. Genel, B. D. Wandelt, *et al.*, "Zooming by in the CARPoolGP lane: new CAMELS-TNG simulations of zoomed-in massive halos," *arXiv e-prints*, arXiv:2403.10609, arXiv:2403.10609, Mar. 2024. DOI: 10.48550/arXiv.2403.10609. arXiv: 2403.10609 [astro-ph.GA].

[412] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, *et al.*, "The Gaia mission," *A&A*, vol. 595, A1, A1, Nov. 2016. DOI: 10.1051/0004-6361/201629272. arXiv: 1609.04153 [astro-ph.IM].

[413] Gaia Collaboration, A. G. A. Brown, A. Vallenari, *et al.*, "Gaia Data Release 2. Summary of the contents and survey properties," *A&A*, vol. 616, A1, A1, Aug. 2018. DOI: 10.1051/0004-6361/201833051. arXiv: 1804.09365 [astro-ph.GA].

[414] A. Helmi, C. Babusiaux, H. H. Koppelman, D. Massari, J. Veljanoski, and A. G. A. Brown, "The merger that led to the formation of the Milky Way's inner stellar halo and thick disk," *Nature*, vol. 563, no. 7729, pp. 85–88, Oct. 2018. DOI: 10.1038/s41586-018-0625-x. arXiv: 1806.06038 [astro-ph.GA].

[415] J. M. D. Kruijssen, J. L. Pfeffer, M. Reina-Campos, R. A. Crain, and N. Bastian, "The formation and assembly history of the Milky Way revealed by its globular cluster population," *MNRAS*, vol. 486, no. 3, pp. 3180–3202, Jul. 2019. DOI: 10.1093/mnras/sty1609. arXiv: 1806.05680 [astro-ph.GA].

[416] G. M. Green, E. Schlafly, C. Zucker, J. S. Speagle, and D. Finkbeiner, "A 3D Dust Map Based on Gaia, Pan-STARRS 1, and 2MASS," *ApJ*, vol. 887, no. 1, 93, p. 93, Dec. 2019. DOI: 10.3847/1538-4357/ab5362. arXiv: 1905.02734 [astro-ph.GA].

[417] T. Cantat-Gaudin, C. Jordi, A. Vallenari, *et al.*, "A Gaia DR2 view of the open cluster population in the Milky Way," *A&A*, vol. 618, A93, A93, Oct. 2018. DOI: 10.1051/0004-6361/201833476. arXiv: 1805.08726 [astro-ph.GA].

[418] A.-C. Eilers, D. W. Hogg, H.-W. Rix, and M. K. Ness, "The Circular Velocity Curve of the Milky Way from 5 to 25 kpc," *ApJ*, vol. 871, no. 1, 120, p. 120, Jan. 2019. DOI: 10.3847/1538-4357/aaf648. arXiv: 1810.09466 [astro-ph.GA].

[419] T. Antoja, A. Helmi, M. Romero-Gómez, *et al.*, "A dynamically young and perturbed Milky Way disk," *Nature*, vol. 561, no. 7723, pp. 360–362, Sep. 2018. DOI: 10.1038/s41586-018-0510-7. arXiv: 1804.10196 [astro-ph.GA].

[420] Gaia Collaboration, A. Vallenari, A. G. A. Brown, *et al.*, "Gaia Data Release 3: Summary of the content and survey properties," *arXiv e-prints*, arXiv:2208.00211, arXiv:2208.00211, Jul. 2022. arXiv: 2208.00211 [astro-ph.GA].

[421] Gaia Collaboration, A. G. A. Brown, A. Vallenari, *et al.*, "Gaia Early Data Release 3. Summary of the contents and survey properties," *A&A*, vol. 649, A1, A1, May 2021. DOI: 10.1051/0004-6361/202039657. arXiv: 2012.01533 [astro-ph.GA].

[422] A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert, "Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-mass Galaxy," *ApJ*, vol. 827, no. 2, L23, p. L23, Aug. 2016. DOI: 10.3847/2041-8205/827/2/L23. arXiv: 1602.05957 [astro-ph.GA].

[423] N. Shipp, N. Panithanpaisal, L. Necib, *et al.*, "Streams on FIRE: Populations of Detectable Stellar Streams in the Milky Way and FIRE," *ApJ*, vol. 949, no. 2, 44, p. 44, Jun. 2023. DOI: 10.3847/1538-4357/acc582.

[424] M. R. Buckley, S. H. Lim, E. Putney, and D. Shih, "Measuring Galactic dark matter through unsupervised machine learning," *MNRAS*, Mar. 2023. DOI: 10.1093/mnras/stad843. arXiv: 2205.01129 [astro-ph.GA].

[425] V. Belokurov, J. L. Sanders, A. Fattahi, M. C. Smith, A. J. Deason, N. W. Evans, and R. J. J. Grand, "The biggest splash," *MNRAS*, vol. 494, no. 3, pp. 3880–3898, May 2020. DOI: 10.1093/mnras/staa876. arXiv: 1909.04679 [astro-ph.GA].

[426] P. F. Hopkins, "A new class of accurate, mesh-free hydrodynamic simulation methods," *MNRAS*, vol. 450, no. 1, pp. 53–110, Jun. 2015. DOI: 10.1093/mnras/stv195. arXiv: 1409.7395 [astro-ph.CO].

[427] S. Sharma, J. Bland-Hawthorn, K. V. Johnston, and J. Binney, "Galaxia: A Code to Generate a Synthetic Survey of the Milky Way," *ApJ*, vol. 730, no. 1, 3, p. 3, Mar. 2011. DOI: 10.1088/0004-637X/730/1/3. arXiv: 1101.3561 [astro-ph.GA].

[428] R. L. Beaton, S. Werner, A. W. Mitschang, *et al.*, "APOGEE-centric Ananke Simulations in a SciServer SQL Database," *Research Notes of the American Astronomical Society*, vol. 6, no. 6, 125, p. 125, Jun. 2022. DOI: 10.3847/2515-5172/ac7808.

[429] DES Collaboration, "The Dark Energy Survey," *ArXiv Astrophysics e-prints*, Oct. 2005. eprint: astro-ph/0510346.

[430] R. E. Sanderson, S. Garrison-Kimmel, A. Wetzel, T. Keung Chan, P. F. Hopkins, D. Kereš, I. Escala, C.-A. Faucher-Giguère, and X. Ma, "Reconciling Observed and Simulated Stellar Halo Masses," *ApJ*, vol. 869, no. 1, 12, p. 12, Dec. 2018. DOI: 10.3847/1538-4357/aaeb33. arXiv: 1712.05808 [astro-ph.GA].

[431] V. P. Debattista, O. A. Gonzalez, R. E. Sanderson, K. El-Badry, S. Garrison-Kimmel, A. Wetzel, C.-A. Faucher-Giguère, and P. F. Hopkins, "Formation, vertex deviation, and age of the Milky Way's bulge: input from a cosmological simulation with a late-forming bar," *MNRAS*, vol. 485, no. 4, pp. 5073–5085, Jun. 2019. DOI: 10.1093/mnras/stz746. arXiv: 1805.12199 [astro-ph.GA].

[432] S. Ansar, S. Pearson, R. E. Sanderson, A. Arora, P. F. Hopkins, A. Wetzel, E. C. Cunningham, and J. Quinn, "Bar formation and destruction in the FIRE-2 simulations," *arXiv e-prints*, arXiv:2309.16811, arXiv:2309.16811, Sep. 2023. DOI: 10.48550/arXiv.2309.16811. arXiv: 2309.16811 [astro-ph.GA].

[433] X. Ma, P. F. Hopkins, A. R. Wetzel, E. N. Kirby, D. Anglés-Alcázar, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert, "The structure and dynamical evolution of the stellar disc of a simulated Milky Way-mass galaxy," *MNRAS*, vol. 467, no. 2, pp. 2430–2444, May 2017. DOI: 10.1093/mnras/stx273. arXiv: 1608.04133 [astro-ph.GA].

[434] K. El-Badry, E. Quataert, A. Wetzel, *et al.*, "Gas kinematics, morphology and angular momentum in the FIRE simulations," *MNRAS*, vol. 473, no. 2, pp. 1930–1955, Jan. 2018. DOI: 10.1093/mnras/stx2482. arXiv: 1705.10321 [astro-ph.GA].

[435] J. Bland-Hawthorn and O. Gerhard, "The Galaxy in Context: Structural, Kinematic, and Integrated Properties," *ARA&A*, vol. 54, pp. 529–596, Sep. 2016. DOI: 10.1146/annurev-astro-081915-023441. arXiv: 1602.07702 [astro-ph.GA].

[436] P. Kroupa, "On the variation of the initial mass function," *MNRAS*, vol. 322, no. 2, pp. 231–246, Apr. 2001. DOI: 10.1046/j.1365-8711.2001.04022.x. arXiv: astro-ph/0009005 [astro-ph].

[437] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969. DOI: 10.1137/1114019. eprint: https://doi.org/10.1137/1114019. URL: https://doi.org/10.1137/1114019.

[438] S. Sharma and M. Steinmetz, "Multidimensional density estimation and phase-space structure of dark matter haloes," *MNRAS*, vol. 373, no. 4, pp. 1293–1307, Dec. 2006. DOI: 10.1111/j.1365-2966.2006.11043.x. arXiv: astro-ph/0507550 [astro-ph].

[439] S. Sharma and K. V. Johnston, "A Group Finding Algorithm for Multidimensional Data Sets," *ApJ*, vol. 703, no. 1, pp. 1061–1077, Sep. 2009. DOI: 10.1088/0004-637X/703/1/1061.

[440] A. C. Robin, X. Luri, C. Reylé, *et al.*, "Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue," *A&A*, vol. 543, A100, A100, Jul. 2012. DOI: 10.1051/0004-6361/201118646. arXiv: 1202.0132 [astro-ph.GA].

[441] Y. Chen, L. Girardi, X. Fu, A. Bressan, B. Aringer, P. Dal Tio, G. Pastorelli, P. Marigo, G. Costa, and X. Zhang, "YBC: a stellar bolometric corrections database with variable extinction coefficients. Application to PARSEC isochrones," *A&A*, vol. 632, A105, A105, Dec. 2019. DOI: 10.1051/0004-6361/201936612. arXiv: 1910.09037 [astro-ph.SR].

[442] R. C. Bohlin, I. Hubeny, and T. Rauch, "New Grids of Pure-hydrogen White Dwarf NLTE Model Atmospheres and the HST/STIS Flux Calibration," *AJ*, vol. 160, no. 1, 21, p. 21, Jul. 2020. DOI: 10.3847/1538-3881/ab94b4. arXiv: 2005.10945 [astro-ph.SR].

[443] N. P. Gentile Fusillo, P. .-. Tremblay, E. Cukanovaite, A. Vorontseva, R. Lallement, M. Hollands, B. T. Gänsicke, K. B. Burdge, J. McCleery, and S. Jordan, "A catalogue of white dwarfs in Gaia EDR3," *MNRAS*, vol. 508, no. 3, pp. 3877–3896, Dec. 2021. DOI: 10.1093/mnras/stab2672. arXiv: 2106.07669 [astro-ph.SR].

[444] H. L. Johnson, "Interstellar Extinction in the Galaxy.," *ApJ*, vol. 141, p. 923, Apr. 1965. DOI: 10.1086/148186.

[445] G. V. Schultz and W. Wiemer, "Interstellar reddening and IR-excesses of O and B stars.," *A&A*, vol. 43, pp. 133–139, Sep. 1975.

[446] D. C. B. Whittet and I. G. van Breda, "Infrared photometry of southern early-type stars.," *MNRAS*, vol. 192, pp. 467–480, Aug. 1980. DOI: 10.1093/mnras/192.3.467.

[447] D. Katz, P. Sartoretti, M. Cropper, *et al.*, "Gaia Data Release 2. Properties and validation of the radial velocities," *A&A*, vol. 622, A205, A205, Feb. 2019. DOI: 10.1051/0004-6361/201833273. arXiv: 1804.09372 [astro-ph.IM].

[448] A. C. R. Thob, R. E. Sanderson, A. P. Eden, F. Nikakhtar, N. Panithanpaisal, N. Garavito-Camargo, and S. Sharma, "Generating synthetic star catalogs from simulated data for next-gen observatories with py-ananke," *arXiv e-prints*, arXiv:2312.02268, arXiv:2312.02268, Dec. 2023. DOI: 10.48550/arXiv.2312.02268. arXiv: 2312.02268 [astro-ph.GA].

[449] J. Binney and G. A. Mamon, "M/L and velocity anisotropy from observations of spherical galaxies, or must M87 have a massive black hole," *MNRAS*, vol. 200, pp. 361–375, Jul. 1982. DOI: 10.1093/mnras/200.2.361.

[450] J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton Series in Astrophysics), Second. Princeton University Press, 2008.

[451] R. Barlow, "Asymmetric statistical errors," in *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, Jun. 2004, pp. 56–59. arXiv: physics/0406120.

[452] M. Lisanti, S. Mishra-Sharma, N. L. Rodd, B. R. Safdi, and R. H. Wechsler, "Mapping Extragalactic Dark Matter Annihilation with Galaxy Surveys: A Systematic Study of Stacked Group Searches," *Phys. Rev. D*, vol. 97, no. 6, p. 063 005, 2018. DOI: 10.1103/PhysRevD.97.063005. arXiv: 1709.00416 [astro-ph.CO].

[453] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[454] F. Nielsen, "On the jensen–shannon symmetrization of distances relying on abstract means," *Entropy*, vol. 21, no. 5, 2019, ISSN: 1099-4300. DOI: 10.3390/e21050485. URL: https://www.mdpi.com/1099-4300/21/5/485.

[455] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press, 1999. URL: http://nlp.stanford.edu/fsnlp/.

[456] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951. DOI: 10.1214/aoms/1177729694. URL: https://doi.org/10.1214/aoms/1177729694.

[457] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, and G. Louppe, *Averting a crisis in simulation-based inference*, 2021. DOI: 10.48550/ARXIV.2110.06581. URL: https://arxiv.org/abs/2110.06581.

[458] B. Dey, D. Zhao, J. A. Newman, B. H. Andrews, R. Izbicki, and A. B. Lee, "Calibrated predictive distributions via diagnostics for conditional coverage," *arXiv preprint arXiv:2205.14568*, 2022.

[459] M. Germain, K. Gregor, I. Murray, and H. Larochelle, "MADE: masked autoencoder for distribution estimation," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, F. R. Bach and D. M. Blei, Eds., ser. JMLR Workshop and Conference Proceedings, vol. 37, 2015, pp. 881–889. URL: http://proceedings.mlr.press/v37/germain15.html.

[460] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics," *arXiv e-prints*, arXiv:1503.03585, arXiv:1503.03585, Mar. 2015. DOI: 10.48550/arXiv.1503.03585. arXiv: 1503.03585 [cs.LG].

[461] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, "Improving Variational Inference with Inverse Autoregressive Flow," *arXiv e-prints*, arXiv:1606.04934, arXiv:1606.04934, Jun. 2016. DOI: 10.48550/arXiv.1606.04934. arXiv: 1606.04934 [cs.LG].

[462] C. Kaae Sønderby, T. Raiko, L. Maaløe, S. Kaae Sønderby, and O. Winther, "Ladder Variational Autoencoders," *arXiv e-prints*, arXiv:1602.02282, arXiv:1602.02282, Feb. 2016. DOI: 10.48550/arXiv.1602.02282. arXiv: 1602.02282 [stat.ML].