

The DORCHA suite: nature, nurture, and the phase space distribution of the Milky Way’s high redshift progenitors today

Sreedhar Balu^{1,2,5*}, Chris Power^{3,5}, Kris Walker^{3,5}, & J. Stuart B. Wyithe^{4,5}

¹Facultad de Físicas, Multidisciplinary Unit for Energy Science, Universidad de Sevilla, 41012, Seville, Spain

²School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

³International Centre for Radio Astronomy Research, M468, University of Western Australia, 35 Stirling Hwy, Perth, WA 6009, Australia

⁴Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

⁵ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Where in the present-day Milky Way should we search for the remnants of its earliest stars? We address this question using the DORCHA (Gaelic for *Dark*; DUR-UH-KHUH) suite: a set of 25 high-resolution, dark-matter-only cosmological zoom-in simulations of Milky Way analogue (MWA) haloes evolved to $z = 0$. Of these, 15 are isolated and the rest are in pairs, similar to the MW and M31. By identifying and tagging the most bound material in high-redshift ($z \geq 5$) progenitor haloes – those likely to host early star formation – we track the present-day phase-space distribution of this ancient component. We find that this material is highly centrally concentrated at $z = 0$, with 90 – 100 per cent residing within $r \lesssim 15 h^{-1}$ kpc. It exhibits steep density profiles ($\rho \propto r^{-4}$), low velocity dispersions ($\sigma_r/\sigma_{\max} \lesssim 0.6$), and radially biased orbits ($\beta \gtrsim 0.5$ for $r \gtrsim 0.1 R_{200}$), consistent with a relaxed, centrally embedded population. These results hold across haloes with diverse formation histories and environments, suggesting that the dynamical signature of early progenitors is robust to later mergers and interactions. Our findings imply that the fossil record of the first generations of stars – including Population III and extremely metal-poor stars – should be sought in the innermost regions of the Milky Way, where they retain distinctive kinematic imprints. While these stellar populations may overlap, we caution that low metallicity does not uniquely identify ancient stars, nor vice versa. The DORCHA suite thus provides a physically motivated baseline for interpreting observations from Galactic Archaeology surveys targeting the bulge and inner halo.

Key words: Galaxy: general – kinematics and dynamics – Cosmology: dark matter – stars: Population III

1 INTRODUCTION

The last several years have witnessed a remarkable advance in astronomers’ capacity to measure the properties of statistical samples of galaxies at high redshifts ($z \gtrsim 5$) and to characterise their larger-scale environment, with important consequences for our understanding of galaxy formation and evolution in the early Universe. This is most strikingly apparent in the results coming from JWST, which have revealed unexpectedly high inferred stellar masses of high redshift massive galaxies (e.g. Boylan-Kolchin 2023; Casey et al. 2024), in apparent tension with theoretical models for early galaxy formation (cf. Stark et al. 2025). Forthcoming direct measurements of the cosmic distribution of neutral atomic hydrogen via the redshifted 21-cm line (e.g. Furlanetto et al. 2006; Morales & Wyithe 2010) by the Square Kilometre Array radio telescope (see, e.g. Mellema et al. 2013) promise a similar transformation in our understanding of the high redshift

Universe. These measurements will reveal the conditions under which early generations of galaxies formed and evolved, and the sources of the ionizing radiation background that led to cosmological reionization (e.g. Barkana & Loeb 2001).

However, we can access a complementary understanding of galaxy formation and evolution in the early Universe by studying its artefacts in the present-day ($z=0$) Universe. Galactic Archaeology (e.g. Freeman & Bland-Hawthorn 2002; Frebel & Norris 2015) targets the Local Group – Milky Way, Andromeda (M31), and their retinue of satellites, globular clusters, and stellar haloesstreams – and dissects the intrinsic properties (e.g. stellar ages and metallicities), spatial structure, kinematics, and inferred dynamics of its constituent systems to characterise their formation and assembly histories. This motivates ongoing Galactic mapping and archaeology surveys such as GALAH (cf. De Silva et al. 2015), Gaia (cf. Gaia Collaboration et al. 2016) and APOGEE (cf. Majewski et al. 2017), and forthcoming surveys and facilities such as 4MOST (cf. De Jong et al. 2019) and LSST (cf. Ivezić et al. 2019).

* E-mail:bsreedhar@us.es

These surveys have access to a wealth of data – positions, kinematics, ages, and metallicities – from which detailed maps of the oldest stars around the Milky Way and in the Local Group have been constructed (cf. Bland-Hawthorn & Gerhard 2016). Age-dating via elemental abundances and stellar astro-seismic signatures reveal that the oldest stars formed within a few hundred million years after the Big Bang (e.g. Freeman & Bland-Hawthorn 2002; Koch et al. 2008; De Silva et al. 2015; Tolstoy et al. 2009). These old stars are found within multiple components – the stellar halo (Carollo et al. 2010; Helmi 2020); the Galactic bulge (Barbuy et al. 2018); dwarf galaxies (Tolstoy et al. 2009; McConnachie et al. 2014); and globular clusters (Pritzl et al. 2005) – and they are observed to trace distinct spatial and kinematic distributions. For example, the old stellar halo population is concentrated in an inner component, overlaps with the Galactic bulge, and it follows a preferentially radial orbital distribution, in contrast to the outer component (cf. Bland-Hawthorn & Gerhard 2016). Although metallicity is often used as a proxy for age, especially in Galactic Archaeology, the two are not strictly equivalent. Extremely metal-poor stars can form after the first generations of star formation, and not all old stars necessarily have low metallicities. This distinction is important for interpreting observational data.

Due to the hierarchical and inside-out growth of dark matter (DM) haloes, we expect the inner regions of the Milky Way to host most of its oldest components (Diemand et al. 2005). Observational campaigns targeting the old (i.e. formed at $z \geq 5$) and metal-poor ($[\text{Fe}/\text{H}] < -2.5$) stars – which serve as luminous tracers of these components – have overlooked the inner regions of the Milky Way because of the challenges disentangling the effects of severe dust obscuration and the dominance of the more numerous metal-rich components. We expect that the number of metal-poor and old stars in the innermost parts of the Milky Way to be negligible fraction of the total number of stars present. A random survey is unlikely to separate the old components (Ness et al. 2013b) from the metal-rich stars in this environment, whereas targeted surveys (e.g. (EMBLA; Howes et al. 2014, 2016), (COMB; Lucey et al. 2019), etc) have indeed detected extremely metal-poor stars within the inner parts of the Milky Way. Recently, the kinematical (Arentsen et al. 2020) and dynamical (Ardern-Arentsen et al. 2024) properties of the low-metallicity stars in the Galaxy's inner bulge were studied using the Pristine Inner Galaxy Survey (PIGS). They report distinct kinematic and dynamical features when separating their samples based on the metallicity.

Given these observed (present-day) properties of the Milky Way's oldest stars, the task of the Galactic Archaeologist is to make inferences about the assembly history of the Milky Way and the conditions under which its high redshift progenitors formed and evolved. It's interesting to ask how sensitive these inferences are to (1) the peculiarities of the Milky Way – such as the relatively isolated nature of the Local Group, and the proximity of Andromeda – and to (2) the assumptions we make about how galaxy formation proceeds at early times.

In this paper, the first of a series, we use cosmological N -body simulations to address the first of these questions. Astrophysical simulations provide a powerful interpretive framework to connect the observable properties of old stellar populations to the assembly history of the Milky Way. For exam-

ple, Bullock & Johnston (2005) used non-cosmological models of disrupting satellite galaxies in a Galactic potential to argue that the stellar halo of the Milky Way most likely grew from the inside-out, with an old, centrally concentrated, stellar component of stars on preferentially radial orbits. Subsequent work using cosmological N -body simulations (e.g. Diemand et al. 2005; Madau et al. 2008; Springel et al. 2008; Gao et al. 2010) showed that such profiles follow naturally if we assume that these old stars are associated with high- σ peaks¹ in the early Universe. The spatial and kinematic distributions of this material in the present day Milky Way will show a number of distinctive features – compared to the overall mass distribution, it will be more centrally concentrated; fall off more steeply with radius; have a lower velocity dispersion; and move on more radial orbits (e.g. Diemand et al. 2005).

The consistency of the present radial mass density profiles of material associated with high redshift high- σ peaks with profiles of the Milky Way's stellar bulge and halo, dwarf galaxies, and globular clusters (cf. Diemand et al. 2005) is highly suggestive that we can map present-day phase space structure to its high-redshift progenitors. Indeed, if old stellar material has been relatively recently accreted by the Milky Way, it may continue to be a coherent substructure in phase space (e.g. Knebe et al. 2005), but this becomes less likely the longer it orbits in the Milky Way's time-varying gravitational potential (e.g. Thomas et al. 2025).

In this paper, we use our new DORCHA² suite of cosmological N -body simulations to study what we can learn about the Milky Way's progenitors in the early Universe and its subsequent assembly history from old stellar populations at $z = 0$. We do this using a statistical sample of 25 Milky Way Analogue (MWAs), selected to have masses and local environments comparable to the Milky Way; this sample consists of 15 isolated systems and 5 pairs. We can track MWA progenitors as early as $z = 25$ and tag the most bound material in the structures capable of supporting atomic hydrogen cooling between $5 \leq z \leq 25$ as potential sites for high redshift star formation; this material provides us with dynamical tracers of the $z = 0$ old stellar populations, which we note may encompass both metal-free and metal-poor stars. Our suite allows us to assess the degree to which assembly history and the close proximity of a massive neighbour is likely to influence the $z = 0$ phase space distribution of the oldest stars in the Milky Way. For example, we expect the Milky Way to have a quiescent merging history, given the relatively early time when its disc was likely assembled (e.g. Bland-Hawthorn & Gerhard 2016); how much of an impact does the plausible range of a MWA's assembly histories have on the old stellar population's phase space? Similarly, how is this phase space influenced if it co-evolves with a pair of galaxies like the Milky Way and Andromeda?

¹ We can characterise overdensities in terms of $\nu\sigma(M, z)$, where ν quantifies the rarity of the overdensity, $\sigma(M, z)$ is the root-mean-square fluctuation of the linearly extrapolated matter density field smoothed with a top-hat filter on a mass scale M , and z is the redshift. A 1- σ peak corresponds to the characteristic mass $M_*(z)$ such that $\sigma(M_*, z) = 1.69$

² From Irish Gaelic, *dorcha* translates to English as *dark*; it is pronounced [dɔr̥χə]; roughly *DUR-uh-khuh*.

We note that the ELVIS suite (cf. Garrison-Kimmel et al. 2014, 48 MWAs with half in pairs); the CATERPILLAR suite (cf. Griffen et al. 2018, 30 MWAs); and the MILKY WAY-EST suite (cf. Buch et al. 2024, 20 MWAs) have adopted a similar approach to us in using cosmological N -body simulations. An important point of difference is that the DORCHA suite is a core part of our upcoming SOLAS suite³ of simulations which will create hydrodynamical galaxy formation counterparts of our N -body haloes. By comparing and contrasting these simulations, we will address the second of the questions posed above (i.e. *How sensitive are our inferences to the assumptions we make about how galaxy formation proceeds at early times?*). The DORCHA suite helps us in developing a heuristic viewpoint that is not tied to specific models of star formation, feedback, and metal enrichment processes.

In the following sections, we outline our methodological approach (§ 2), including a description of our simulations and analysis tools; we present our results (§ 3), including a visual impression of the DORCHA suite, spherically averaged profiles, phase space structure, and orbital distributions; and we provide a summary of our conclusions (§ 4). Unless otherwise specified, we use comoving units.

2 METHODOLOGY

In this section, we present details on how we have run our simulations suite and our approach to halo finding and merger tree construction (§ 2.1); we describe how we selected the MWA candidates for resimulation (§ 2.2); and we outline our approach to identifying material associated with high redshift haloes that might be plausible sites of early star formation (§ 2.4). This is the material whose spatial and kinematic properties we quantify in the next section.

2.1 Simulations

2.1.1 Cosmological Parameters

We assume the Planck Collaboration et al. (2020) cosmological parameters, specifically the TT,TE,EE+lowE+lensing+BAO column of their Table 2. The key parameters are, respectively, the matter, dark energy, and baryon density parameters; the present-day Hubble parameter, the mass variance normalisation; and the spectral index, with values of $(\Omega_m, \Omega_\Lambda, \Omega_b, H_0, \sigma_8, n_s) = (0.3121, 0.6879, 0.0491, 67.51, 0.815, 0.9653)$.

2.1.2 Simulation Code

We use AREPO (Springel 2010; Weinberger et al. 2020) to run our cosmological N -body parent and zoom simulations. AREPO employs a hybrid Tree-PM approach to estimate the gravitational force, splitting the calculation into a long-range component, which is solved via a particle mesh (PM) approach, and a short-range component, which is solved via a hierarchical oct-tree approach (cf. Barnes & Hut 1986; Springel 2005). Each simulation is run from a starting redshift of $z = 99$ to $z = 0$, and we use MUSIC (Hahn & Abel 2011) to generate our initial conditions (ICs).

³ The Gaelic Irish word *solas* means *light*.

2.1.3 The Simulation Suite

Our parent simulation, DORCHA_00, is a uniform resolution periodic volume of side length $L=100 h^{-1}\text{Mpc}$ containing $N=512^3$ particles (i.e. MUSIC input parameter `levelmin=9`), each of mass $m_p \simeq 6.45 \times 10^8 h^{-1}\text{M}_\odot$. We use a fixed comoving softening length of $\epsilon_f = 0.04 \times \ell$ (Ludlow et al. 2019), where $\ell = L/N$ is the mean inter-particle separation. We recast ℓ in the form,

$$\ell = (m_{\text{part}}/\Omega_m \rho_{\text{crit}})^{1/3}, \quad (1)$$

where m_{part} is the particle mass, Ω_m is the matter density parameter, and ρ_{crit} is the critical density; this is convenient when evaluating softenings for zoom simulations. This gives $\epsilon_f = 10 h^{-1}\text{kpc}$.

From the DORCHA_00 halo catalogue at $z=0$, we build a list of candidate haloes for resimulation based on a set of criteria, which we set out in § 2.2, to select the hosts of Milky Way Analogues (MWAs). Our candidates consist of 15 isolated systems (DORCHA_01-15) and 10 in pairs (DORCHA_16-20). We select all the particles within some threshold radius r_{cut} at $z = 0$, and use their positions at $z = 99$ (in the ICs) to define the spatial extent of the Lagrangian volume to be resimulated. For our isolated MWAs, $r_{\text{cut}} = 4 \times R_{200\text{crit}}$ where $R_{200\text{crit}}$ is the virial radius, while $r_{\text{cut}} = 1.5 h^{-1}\text{Mpc}$ relative to their centre of mass for our paired MWAs. In all cases, we add a buffer of $0.5 h^{-1}\text{Mpc}$ along each Cartesian dimension at $z = 99$. The ICs of the MWA zoom simulations have an increased mass resolution of 4096 times relative to DORCHA_00 (i.e. MUSIC input parameter `levelmin=13`). This corresponds to a particle mass in the zoom region of $1.57 \times 10^5 h^{-1}\text{M}_\odot$ for the DORCHA suite of zoom simulations. The associated gravitational force softening of particles in the zoom region is $\epsilon_f = 0.5 h^{-1}\text{kpc}$.

2.1.4 Halo Finding & Merger Trees

We use the version of SUBFIND-HBT in GADGET-4 (Springel et al. 2021) code to construct our halo catalogues. Haloes correspond to Friends-of-Friends (hereafter FoF) groups assuming a linking length of $b = 0.2 \times \ell$. We set a lower limit of 32 particles for FoF groups, which gives us a minimum halo mass of $M_{h,\text{min}}^{\text{parent}} \simeq 2.06 \times 10^{10} h^{-1}\text{M}_\odot$ in the DORCHA_00 parent simulation, and $5.04 \times 10^6 h^{-1}\text{M}_\odot$ in the DORCHA_01-20 zoom simulations.

For the purposes of this paper, we treat FoF groups (GROUPS in SUBFIND-HBT) as haloes, which includes both gravitationally bound and unbound material associated with the halo. Our analysis can be applied equally to the first-ranked (central) subhalo within the FoF group (from SUBHALOES in SUBFIND-HBT), which includes only gravitationally bound material associated with the smooth background halo. However, our inclusive definition means that we maximise the volume within which we might expect to find the remnants of progenitor haloes that hosted high redshift star formation.

Note that in our analysis we use the values of the virial mass, M_{200} , and virial radius, R_{200} computed by SUBFIND-HBT; these values are defined such that the mean enclosed density within R_{200} at redshift z is 200 times the critical density, $\rho_{\text{crit}}(z) = 3H^2(z)/8\pi G$, where $H(z)$ is the Hubble parameter and G is the gravitational constant.

SUBFIND-HBT allows us to reliably track haloes and sub-

Table 1. Details of all 25 DORCHA haloes. DORCHA_01-15 are isolated systems and DORCHA_15-20 comprise of paired systems. M_{200} , R_{200} , and V_{200} correspond to the virial masses, radii, and circular velocities of the haloes at $z=0$; N_{subs} is the number of subhaloes associated with the halo at $z=0$; z_{init} is the redshift at which the main progenitor of the halo was first identified in the SUBFIND-HBT catalogues; z_{10} which is the redshift at which 10 per cent of the halo's $z=0$ mass was in place ; and $z_{\text{form}} \equiv z_{50}$ is the formation redshift, which by convention is when 50 per cent of the main branch mass was in place. We also highlight which haloes can be considered early or late forming (kind).

Halo	M_{200} ($\times 10^{12} h^{-1} M_{\odot}$)	R_{200} ($h^{-1} kpc$)	V_{200} (km s $^{-1}$)	N_{subs} ($z = 0$)	z_{init}	z_{10}	z_{50} ($\equiv z_{\text{form}}$)	kind
DORCHA_01	1.47	185.02	185.0	2867	28.96	3.23	0.28	late
DORCHA_02	1.46	184.46	184.45	2660	27.97	4.75	2.22	early
DORCHA_03	1.22	173.68	173.67	2559	28.96	3.7	1.52	—
DORCHA_04	1.31	177.99	177.97	2194	22.82	2.71	1.06	—
DORCHA_05	1.08	166.75	166.72	2092	23.58	3.88	1.69	early
DORCHA_06	1.16	171.04	171.03	1872	26.17	4.21	1.93	early
DORCHA_07	1.06	165.7	165.68	2261	22.06	3.79	1.56	—
DORCHA_08	0.83	153.0	152.98	1737	13.56	2.33	1.17	—
DORCHA_09	0.72	146.04	146.02	1469	25.3	4.12	1.38	—
DORCHA_10	1.46	184.49	184.47	3135	22.82	3.93	1.72	early
DORCHA_11	1.03	164.45	164.44	1661	22.06	3.85	1.41	—
DORCHA_12	0.95	160.08	160.07	1837	27.07	3.97	1.36	—
DORCHA_13	0.73	146.7	146.68	1342	28.96	3.69	1.37	—
DORCHA_14	0.86	154.5	154.48	1353	22.06	4.26	1.28	—
DORCHA_15	0.65	140.62	140.6	1759	22.82	3.16	1.32	—
DORCHA_16A	1.57	189.1	189.07	3207	26.14	3.35	1.33	—
DORCHA_16B	1.57	188.89	188.87	3001	22.82	3.73	1.23	—
DORCHA_17A	1.08	166.72	166.7	2301	25.3	2.66	0.89	late
DORCHA_17B	0.75	147.42	147.41	2630	23.58	2.86	1.13	—
DORCHA_18A	1.07	166.32	166.3	2521	27.07	3.4	0.63	late
DORCHA_18B	0.89	156.7	156.68	1702	24.43	3.48	1.48	—
DORCHA_19A	1.01	163.2	163.18	1867	28.96	2.86	2.21	early
DORCHA_19B	0.93	158.71	158.7	2051	23.61	3.3	0.66	late
DORCHA_20A	0.95	160.07	160.06	2358	28.96	3.09	1.0	late
DORCHA_20B	0.91	157.31	157.29	2087	22.82	3.95	1.33	—

haloes across snapshots. This takes advantage of the decomposition of FoF haloes into their constituent central subhalo plus associated subhaloes – material associated with a given subhalo is tracked forward in time across snapshots explicitly as part of the halo finding process, from the point at which

it was first identified as a distinct dynamically self-consistent structure. This reduces the ambiguity when linking haloes and subhaloes across snapshots and consequently produces more reliable merger trees (cf. [Chandro-Gomez et al. 2025](#)).

2.1.5 Snapshot Cadence

We use the same snapshot cadence for both parent and zoom simulations – 123 snapshots between $z = 30$ and $z = 0$.

- There are 50 snapshots evenly spaced in $\log(a)$, the natural logarithm of the scalefactor, between $z = 30$ and 5; this corresponds to a snapshot cadence with a minimum(maximum) temporal separation between consecutive snapshots of $\sim 5(\sim 57)$ Myr. This ensures that the temporal separation between two consecutive snapshots is less than the typical dynamical times of the haloes; this is important for accurate halo-finding and merger-tree constructions, especially for merging and accreting haloes, as is common at high redshifts.
- There are 73 snapshots between $z = 5$ and 0. This cadence is tailored to accurately track the orbits of individual particles at late times (see § 3.3).

2.2 Milky Way analogue selection

We wish to study haloes whose properties are broadly consistent with those we might expect for the Milky Way’s host halo at $z = 0$. To obtain these Milky Way Analogue (MWAs), we enforce a *mass criterion* and a *neighbourhood criterion*. Our mass criterion targets haloes with masses in the range $1.04 \times 10^{12} h^{-1} M_\odot \leq M_h \leq 1.95 \times 10^{12} h^{-1} M_\odot$ (see Peñarubia et al. 2015; Cullingham et al. 2019, etc), which yields a sample of 1736 haloes. Our neighbourhood criterion selects isolated and paired MWAs.

- To obtain our isolated MWAs, we require that there are no other haloes within $r_{\text{iso,cut}} = 1.4 h^{-1} \text{Mpc}$ of the mass selected sample of 1736 haloes. There are 17 such haloes in DORCHA_00, which we visually inspect; from these, we choose DORCHA_01–15.
- For our paired MWAs (DORCHA_16–20), we first select haloes that have one equally massive neighbour within $1 h^{-1} \text{Mpc}$ from the mass selected sample of 1736 haloes, and we then retain only those pairs that have no haloes within $r_{\text{iso,cut}} = 0.5 h^{-1} \text{Mpc}$. This yields 21 systems, from which we select 5 pairs.

Figures 1 and 2 provide a visual impression of the parent simulation and the immediate environments of the isolated and pair MWAs respectively. Here we project the mass density along the z -direction onto the x - y plane, such that lighter(darker) hue indicate higher(lower) dark matter density. The large top-left subplot of Figure 1 shows the entire (i.e. $100 h^{-1} \text{Mpc} \times 100 h^{-1} \text{Mpc}$) DORCHA_00 (parent) volume, with blue and pink circles indicating the locations of the isolated and paired MWAs; the remaining subplots show regions of $200 h^{-1} \text{kpc} \times 200 h^{-1} \text{kpc}$ centred on the isolated MWAs, DORCHAS (#01 to #15). Figure 2 shows regions of $1 h^{-1} \text{Mpc} \times 1 h^{-1} \text{Mpc}$ centred on the centre-of-mass of the paired MWAs, DORCHAS (#16 to #20).

In Figures 3 and 4 we characterise the mass assembly histories for our sample of DORCHAS. In Figure 3, the black(brown) curve shows the mean(median) mass growth, while the red dashed curves bracket the interdecile range (i.e. 10th to 90th percentiles); this indicates that our haloes have relatively quiescent assembly histories, which is a consequence of our choice of neighbourhood criterion that means the DORCHAS

are relatively isolated systems. In Figure 4, we show the distribution of the DORCHAS’ formation redshifts, z_{form} ; the mean(median) of the distribution is 1.325(1.33). This reveals that the DORCHAS have a wide variety of assembly histories, with some forming as early as $z_{\text{form}} = 2.22$ (DORCHA_02) and others (DORCHA_01) forming as late as $z_{\text{form}} = 0.28$.

Table 1 lists properties of all 25 DORCHA haloes. Note that among the paired DORCHAS, we follow the naming convention such that the more(less) massive one is labelled A(B). Apart from standard summary statistics to characterise the haloes’ virial properties at $z=0$ (M_{200} , R_{200} , $V_{200} = (GM_{200}/R_{200})^{1/2}$), we also provide information about the number of subhaloes associated with them (N_{subs}); when they was first identified in the SUBFIND-HBT catalogues (z_{init}); and when 10 and 50 per cent of their $z=0$ mass was first in place (z_{10} and z_{50}). The “kind” keyword indicates whether a halo is considered to be early or late forming.

2.3 Splitting by formation time

To better understand the impact and explore if there is any correlation between the particles and the formation time of the haloes, we further categorize the DORCHA haloes into late- and early-forming ones. As mentioned earlier, the formation redshift is defined as the redshift at which 50 per cent of the halo mass at $z = 0$ has been formed. We use this z_{form} (second-to-last column of Table 1 and Figure 4) as proxy for the amount of mergers/accretion that the halo has gone through: the halo has a higher chance of having undergone mergers before its z_{form} .

The mean(median) formation redshift for our haloes is 1.325(1.33). To create our lists of late- and early-DORCHAS, we select 5 haloes each furthest from above and below the mean z_{form} . Thus selected, the late-forming haloes are DORCHA_01 ($z_{\text{form}} = 0.28$), DORCHA_18A ($z_{\text{form}} = 0.63$), DORCHA_19B ($z_{\text{form}} = 0.66$), DORCHA_17A ($z_{\text{form}} = 0.89$), and DORCHA_20A ($z_{\text{form}} = 1.0$). The early-forming haloes are DORCHA_02 ($z_{\text{form}} = 2.22$), DORCHA_19A ($z_{\text{form}} = 2.21$), DORCHA_06 ($z_{\text{form}} = 1.93$), DORCHA_10 ($z_{\text{form}} = 1.72$), and DORCHA_05 ($z_{\text{form}} = 1.69$).

2.4 Tagging High Redshift Remnants

We are interested in picking out regions in phase space where we might expect to find the remnants of star formation in the high-redshift progenitors of the MWAs we expect to reside within our DORCHA haloes. For each DORCHA halo, we identify and tag the particles that are associated with its progenitor haloes at redshifts $z \geq 5$. We do this by creating lists of tagged particles at $z \in [25, 20, 15, 10, 8, 5]$. Where a halo’s progenitors are not resolved by $z_{\text{init}} = 25$, we start tagging at the first redshift at which a progenitor is recovered (sixth column of Table 1). At each of these redshifts, we tag the most-bound 10 per cent⁴ of particles in each of the progenitor haloes – this ensures that we follow the cores of these progenitors. The minimum halo mass of the DORCHA suite – $M_{h,\min} \simeq 5.04 \times 10^6 h^{-1} M_\odot$ – is comparable to the halo

⁴ We have repeated our analyses with 5 and 20 per cent cuts. The precise value of this per cent threshold does not impact our conclusions.

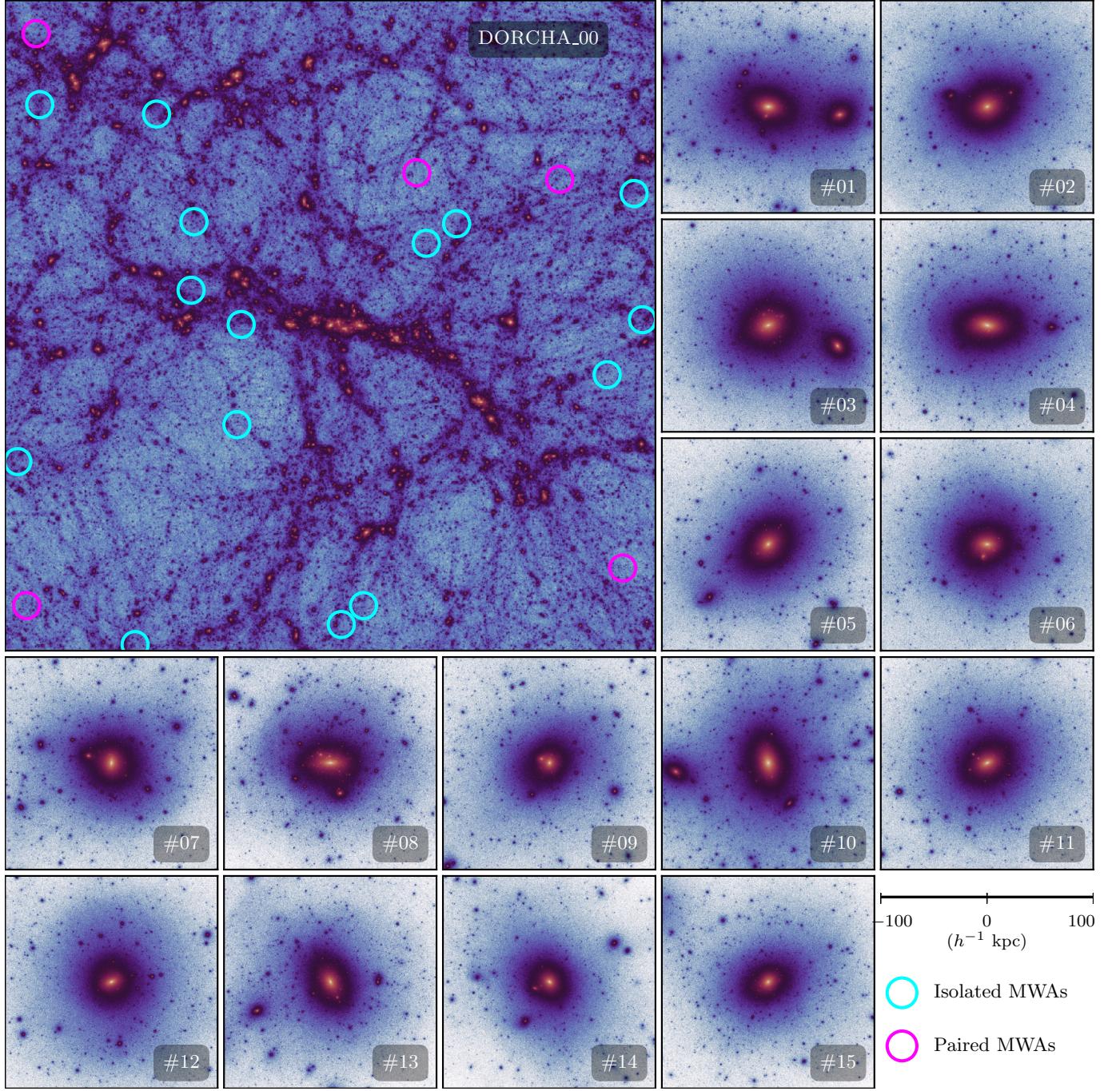


Figure 1. Isolated DORCHA haloes along with the parent DORCHA_00: The subplots show the total projection along the z -direction for our simulations. The top-left larger subplot shows $100 h^{-1}\text{Mpc} \times 100 h^{-1}\text{Mpc}$ projection of our parent DORCHA_00 simulation, with the locations of the isolated and paired Milky Way Analogue (MWAs) - our DORCHA haloes - indicated by the blue and pink circles respectively. The smaller subplots show $200 h^{-1}\text{kpc} \times 200 h^{-1}\text{kpc}$ cutouts around the DORCHA haloes from the DORCHA_01–15 simulations. These figures illustrate the diverse nature, in number as well as spatial distribution, of the subhaloes within the DORCHA suite.

mass at which cooling becomes efficient at $z \sim 40–20$ (cf. Nebrin et al. 2023), and so we might expect these cores to trace where star formation is likely to occur. In this way, we can identify the material in a $z = 0$ descendant that was once in the cores of its progenitors at $z \geq 5$, which could trace where we might expect to find the oldest stars and their remnants.

For comparison, we also tag particles based on when they

first formed part of a halo – "old" material was in place in progenitor haloes before $z = 10$, while "new" material was accreted onto the main branch of the halo after $z = 5$.

We remind the reader that the DORCHA suite is, by construction, dark-matter-only, and so we caution that our predictions miss the potential impact of baryon physics, even though our prescription for assigning the sites of star forma-

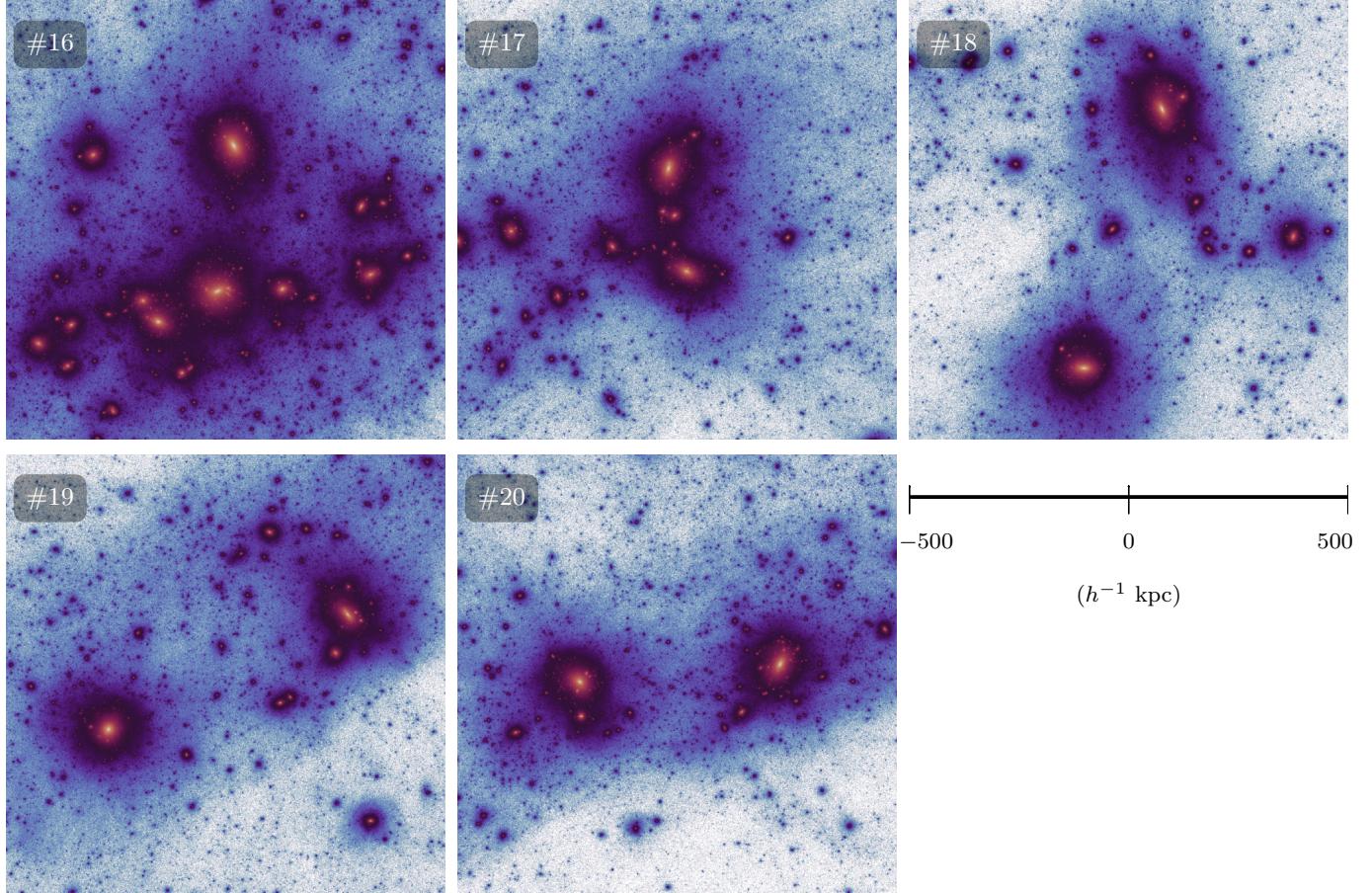


Figure 2. Paired DORCHA haloes: Similar to Figure 1, but for the paired DORCHA simulations (DORCHA_15-20). The cutouts show $1 h^{-1} \text{Mpc} \times 1 h^{-1} \text{Mpc}$ centred at the centre-of-mass of the two haloes in the pair.

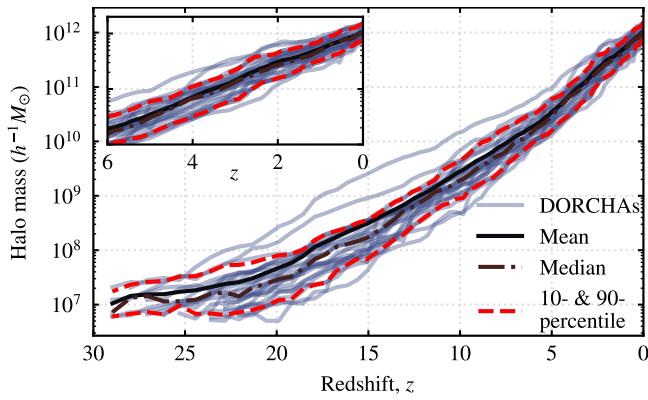


Figure 3. Halo growth histories: We show the mean(black), median(brown), 10-and 90-percentile(red), and individual(light-grey) growth histories of all 25 DORCHA haloes. The inset on the top-left zooms in onto $z \leq 6$. *Takeaway:* Almost all of the haloes have progenitors that are resolved by $z \sim 25$ and have formation histories that are relatively smooth.

tion are rooted in physically sound considerations. Having said this, the conclusions we draw from the DORCHA suite will serve as a baseline to compare against for our future simulations, which will explicitly model hydrodynamical galaxy

formation. We emphasise that while our tagged particles are intended to trace the sites of early star formation, we do not explicitly distinguish between long-lived Population III stars and their extremely metal-poor (but not metal-free) descendants, although these populations may have different formation histories and survival rates. See §3.4 for relevant discussion.

2.5 Measurement of Radial Profiles

Radial profiles can be affected by the particular choice of the centre of the halo (cf. Navarro et al. 2010). When constructing spherically averaged profiles, we select particles within $r/R_{200} \leq 3$ of the halo centre, which we define as the position of the minimum gravitational potential recovered by SUBFIND-HBT. These particles are then binned into 50 radial bins spaced evenly in logarithm between $10^{-3} \leq r/R_{200} \leq 3$ for the radial profiles. The radius of each shell is computed from the mean radii of the particles within the shell.

The mass density profile ($\rho(r)$) is obtained by summing over the mass of particles in the shell and dividing by the shell volume. The mean radial velocity (\bar{v}_r) and velocity dispersion

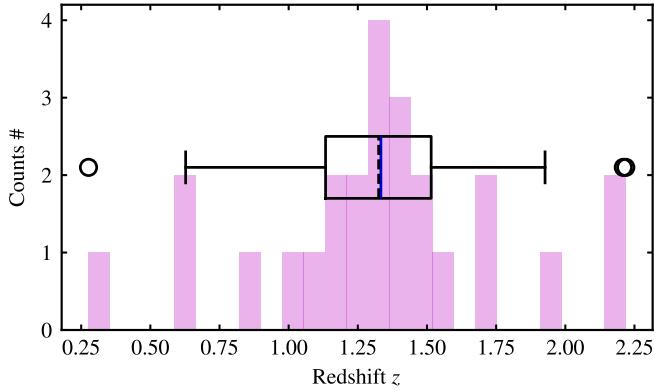


Figure 4. Formation redshift, $z_{\text{form}} \equiv z_{50}$: Distribution of the redshifts at which 50 per cent of the present day mass of the DORCHA haloes were accreted. See also the last column of Table 1. The box-plot encompasses the $1-\sigma$ values. Indicated also are the mean ($=1.325$) in solid and median ($=1.33$) in dashed vertical lines. *Takeaway:* The DORCHA haloes exhibit a variety of formation redshifts, enabling us to cleanly separate out the impact of the surroundings (early- and late-forming kinds) and mergers.

(σ_r) profiles are calculated from

$$\bar{v}_r = N_{\text{shell}}^{-1} \sum_i \frac{\vec{r}_i}{r_i} \cdot (\vec{v}_i - \vec{v}_{\text{cm}}), \quad (2)$$

which gives the mean radial velocity within a shell, with \vec{r}_i being the position vector of the i th particle and \vec{v}_{cm} the centre-of-mass velocity of the halo;

$$\overline{v_r^2} = N_{\text{shell}}^{-1} \sum_i \left[\frac{\vec{r}_i}{r_i} \cdot (\vec{v}_i - \vec{v}_{\text{cm}}) \right]^2, \quad (3)$$

which gives the mean square radial velocity within a shell; and

$$\sigma_r = [\overline{v_r^2} - \bar{v}_r^2]^{1/2}, \quad (4)$$

which gives the corresponding radial velocity dispersion in the shell. We compute the total velocity dispersion as,

$$\sigma_{3D}^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = \sigma_r^2 + \sigma_{\tan}^2, \quad (5)$$

which allow us to obtain the tangential velocity dispersion σ_{\tan} from the radial (σ_r) and Cartesian components of the velocity dispersion ($\sigma_{x/y/z}$). This allows us to compute the velocity anisotropy,

$$\beta(r) = 1 - \frac{1}{2} \frac{\sigma_{\tan}^2}{\sigma_r^2}, \quad (6)$$

which characterises how radial orbits are; for example, $\beta \rightarrow 1$ indicates preferentially radial orbits, while $\beta \rightarrow 0$ will be preferentially isotropic. In addition, we estimate the mean rotational velocity within each radial shell as,

$$\bar{v}_{\text{rot}} = J_{\text{shell}} M_{\text{shell}}^{-1} R_{\text{shell}}^{-1} \quad (7)$$

where $J_{\text{shell}} \equiv |\vec{J}_{\text{shell}}|$ is the magnitude of the angular momentum vector within the shell relative to the centre of mass velocity of the halo, and M_{shell} and R_{shell} are the mass and radius of the shell respectively.

3 RESULTS

In the following subsections, for our sample of $z = 0$ DORCHA haloes, we examine the spherically averaged profiles, assessing their mass density and kinematic profiles (§ 3.1). We also study the phase space distributions (§ 3.2) and the orbital distributions and pericentres (§ 3.3) of material within R_{200} . Our aim is to isolate regions in phase space where we are most likely to find the material from high redshift progenitors in MWAs, and to assess the extent to which these results are likely to be affected by a MWA’s assembly history and galaxy formation processes in its central regions.

3.1 Spherically Averaged Profiles of High Redshift Remnants

In Figures 5, 6, and 7 we show spherically averaged radial profiles for the DORCHA haloes. In each figure, we show (from top to bottom),

- (i) the radial mass density profile in the form $\rho(r) \propto r^2$, normalised by $\rho_{\text{crit}} \times r_{200}^2$;
- (ii) the radial velocity, $\bar{v}_r(r)$, in units of km s^{-1} ;
- (iii) the radial velocity dispersion, $\sigma_r(r)$, normalised by its peak value, $\sigma_{r,\text{max}}$;
- (iv) the velocity anisotropy, $\beta(r)$; and
- (v) the mean rotational velocity, \bar{v}_{rot} , in units of km s^{-1} .

The left column shows the results for all of the material within each of the haloes at $z=0$; the middle column focuses on only those particles that were in the high redshift ($z \geq 5$) progenitors of the $z=0$ halo, and shows profiles for $5 \leq z \leq z_i = 25$; and the right column shows the results for material that were in place at $z \geq 10$ ("old") and material accreted at $z \leq 5$ ("new"). In the first column we show the curves for all the DORCHAS (in light-green) while for the other columns we only show the average trends – mean(median) trend in solid(dotted) curves.

3.1.1 The full DORCHA sample

Mass Density: The top left panel of Figure 5 shows that the mass density profiles of each of the 25 DORCHA haloes (green curves) exhibit little variation with respect to each other out to R_{200} , irrespective of whether they are isolated or paired; consistent with Garrison-Kimmel et al. (2014). Relative to the median profile (dark solid curve), which overlaps the mean profile (dark dashed curve), we find that the measured $\rho(r)$ differs by approximately no more than 30 per cent within $\sim R_{200}$, and the most significant deviations arise in the outskirts.

In the top middle panel, we focus on those particles that are associated with high redshift progenitors. The 6 curves correspond to the present day median (solid) and mean (dashed) profiles of particles that were already in progenitors that formed by $z = (5, 8, 10, 15, 20, z_i)$. As we expect, the density profiles of material associated with earlier forming progenitors have a lower amplitude (because they contain less mass); they are more centrally concentrated (the radius, r_{-2} , at which the peak of the ρr^2 profile occurs increases by approximately 0.5 dex for material associated with $z = z_i$ and $z = 5$ progenitors); and they truncate at progressively smaller

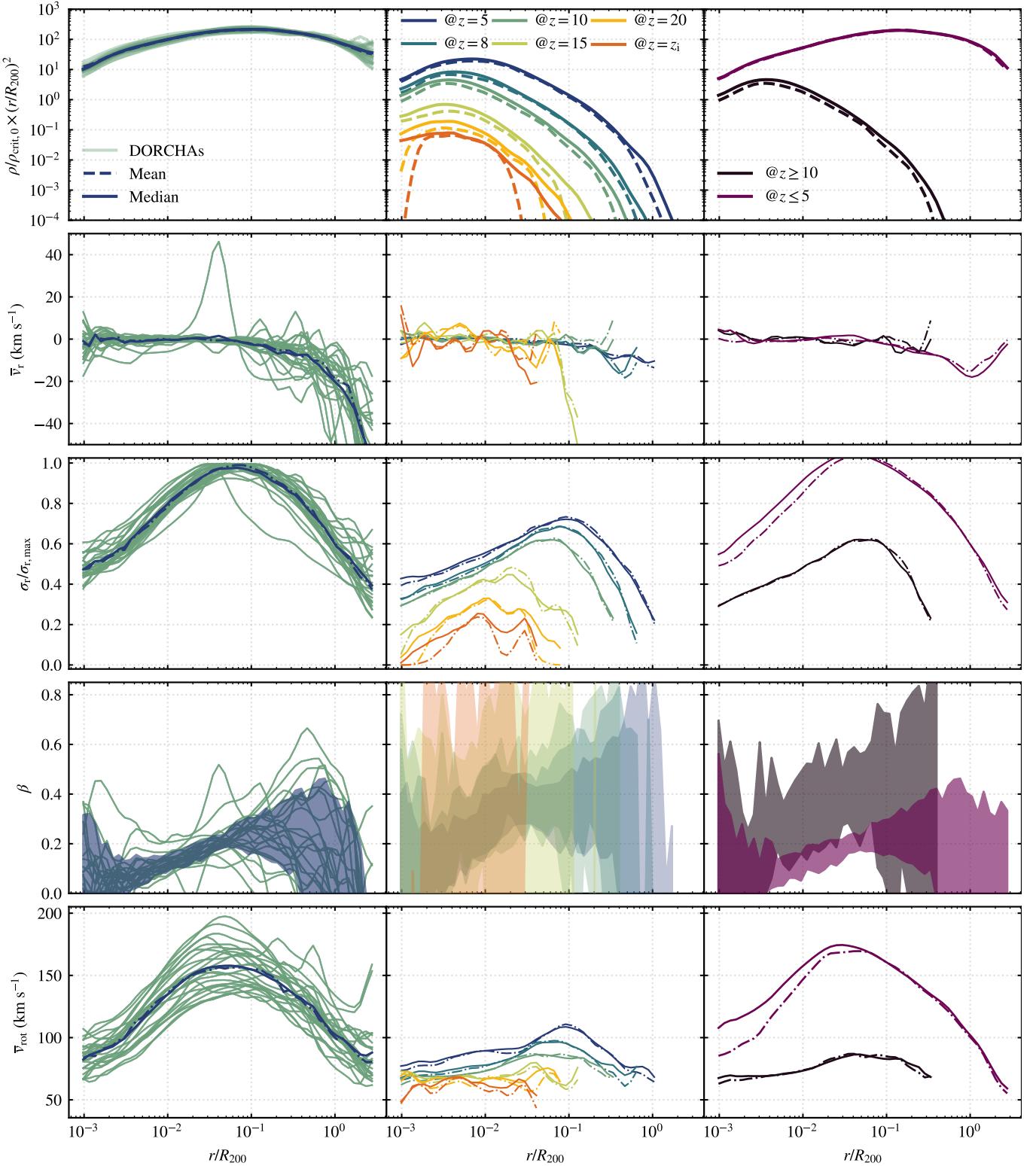


Figure 5. Radial profiles for all DORCHAs: The rows from top to bottoms shows, respectively, the density profile (ρ), mean radial velocity (\bar{v}_r), radial velocity dispersion (σ_r), velocity anisotropy (β), and mean rotational velocity (\bar{v}_{rot}). From left to right the columns are for, respectively, all the particles that comprise the halo, the particles that were tagged at high redshifts in different colours, and the particle list for old ($z \geq 10$) and new ($z \leq 5$) particles. Mean(median) trend is shown with solid(dashed) curves. The shaded region for the β -profiles show the $1-\sigma$ variation. The particles are binned into 50 radial bins evenly spaced in logarithm between $10^{-3} \leq r/R_{200} \leq 3$.

Takeaway: The radial profiles in the 2nd-column shows that most of the high- z progenitors are now part of the main subhalo at $z = 0$. The earlier a particular set of particles were tagged, the more centrally concentrated they are at $z = 0$.

radii. Interestingly, these profiles have a similar shape, with an outer slope that can be approximated as $\rho \propto r^{-4}$.

In the top right panel, we separate particles into those that were in progenitors prior to $z = 10$ (the old component) and those accreted since $z = 5$ (the new component); as before, the mean and medians are shown by the dashed and solid curves. This emphasises how much more centrally concentrated the old component is relative to the young component ($r_{-2}^{\text{old}}/R_{200} \simeq 0.003$ compared to $r_{-2}^{\text{new}}/R_{200} \simeq 0.1$) and how much more rapidly the mass density of the old component drops off relative to the new ($\rho^{\text{old}} \sim 10^{-1} \rho^{\text{new}}$ at $0.003R_{200}$ to $\rho^{\text{old}} \sim 10^{-3} \rho^{\text{new}}$ at $0.1R_{200}$).

Kinematics: The middle three rows of Figure 5 show the mean radial velocity (\bar{v}_r); radial velocity dispersion (σ_r); and velocity anisotropy (β) profiles.

In the left panels, we show the kinematic profiles for each of the 25 DORCHA haloes in green with the median and mean profiles overlaid by the heavy navy solid and dashed curves. The median behaviour is for $\bar{v}_r \simeq 0$ within $\sim 0.8 R_{200}$, before declining in the outskirts, at $\sim R_{200}$, as a result of infalling material. Individual haloes show local variations of $\pm 20 \text{ km s}^{-1}$ in the outskirts, while one (DORCHA_17B) has a nearby companion which manifests as a spike at $10^{-2} \leq r/R_{200} \leq 10^{-1}$. The median velocity dispersion peaks at $\simeq 0.06 R_{200}$; individual profiles typically fall within $\pm 5\%$, with the exception of DORCHA_17A. The velocity anisotropy increases from $\beta \simeq 0$ (isotropic) in the centre to $\beta \simeq 0.3$ (preferentially radial) at R_{200} . The shaded region indicates the 10-to-90 percentile range, which shows that individual haloes have velocity distribution in the outskirts that vary from isotropic ($\beta \simeq 0$) to strongly radial ($\beta \simeq 0.5$).

In the middle panels, we plot the median and mean kinematic profiles for the material associated with the high redshift progenitors identified between $z = 5$ and $z = z_i$. The median behaviour is for $\bar{v}_r \simeq 0$, which indicates that the material is in dynamical equilibrium; there is no evidence for a decline in $\bar{v}_r \simeq 0$ in the outskirts, which implies that all of the progenitor material has been assembled in the main body of the halo for several dynamical times. The median velocity dispersion of the $z = 5$ component peaks at $\simeq 0.06 R_{200}$; while its shape is preserved, the magnitude and location decreases as we probe material from higher redshift components (c.f. Navarro et al. 2010). This indicates that the material that assembled earliest in progenitors that collapsed at $z \geq 10$, which is most centrally concentrated, is dynamically colder – as we would expect for material in the innermost parts of the gravitational potential. The velocity anisotropy profiles for each of the components have similar shapes, showing a characteristic increase from $\beta \simeq 0$ (isotropic) in the centre to $\beta \simeq 0.5$ (preferentially radial) at R_{200} . As the shaded regions reveal, there is a larger variation in β but the general trend is for more strongly radial orbits for the higher redshift components.

In the right panels, we plot the median and mean kinematic profiles for the material associated with the old and new components ($z \geq 10$ and $z \leq 5$ respectively). The trends are similar to those in the middle panels; the median $\bar{v}_r \simeq 0$, which indicates that the material is in dynamical equilibrium, while the median velocity dispersions of the two component peaks at $\simeq 0.06 R_{200}$, differing only in magnitude. The inner slope of the velocity dispersion profile of the old component is shallower than that of the new component, as we would

expect for material orbiting predominantly in the innermost parts of the potential for many dynamical times (see § 3.3 for more discussion). This is reflected in the differences between the velocity anisotropy profiles; both show an increase with radius holds, but the old component is more strongly preferentially radial, and the pronounced turnover at $0.1R_{200}$ is as we would expect for material that is close to apocentre.

Angular Momentum: In the bottom row of Figure 5, we show the mean rotational velocity, \bar{v}_{rot} . The trend shows the expected one for the dark matter haloes. The \bar{v}_{rot} rises with radius in the inner regions of the halo due to the steep mass profile. Further away from the centre, the \bar{v}_{rot} declines as the density decreases. We note that, from the third column, the older material have a comparatively flat \bar{v}_{rot} profile.

General comments: There are few markedly different features between the old and the new components that could help one distinguish between them observationally. If, within a given radial, we can identify the subset of stars with low radial and rotational velocities when compared to the full population in that bin, then there is a high chance that they belonged to one of the high- z ($z \geq 10$) progenitors of the MW. Our results are in qualitative agreement with (Arentsen et al. 2020), who studied the kinematics of the metal-poor stars in the Galactic bulge as a function of metallicity and reported a decreasing rotational velocity with decreasing metallicity for the metal-poor stars. This is interesting because it suggests that the impact of galaxy assembly and the presence of a stellar bar on the inner potential will affect our results quantitatively but not qualitatively. Of course, we can verify this in our forthcoming simulations that model hydrodynamical galaxy formation.

3.1.2 The early- and late-forming DORCHA sample

In Figure 3 we showed that the DORCHA haloes have relatively quiescent histories with no evidence for major mergers in their recent assembly history. However, for each halo we have a formation redshift, z_{form} – the redshift at which half of the mass of the main progenitor of the $z=0$ halo is first in place (see § 2.3) – which we show in Figure 4. Based on this information, we select the five earliest and five latest forming haloes (i.e. those with the largest and smallest z_{form} values respectively) from the DORCHA sample and repeat our analysis of the previous subsection in Figures 6 and 7.

Despite the difference in the typical formation redshift of the earliest-forming haloes ($z_{\text{form}} \simeq 2$) compared to the latest-forming haloes ($z_{\text{form}} \simeq 0.6$), we find remarkably little variation in the mean and median profiles. This is surprising because the latest-forming haloes are preferentially those in pairs, where we might expect the presence of a massive neighbour to have an influence. The mass density profiles are consistent with one another – the trends we noted in the previous section remain regardless of how we split by z_{form} – while differences in the kinematics profiles can be explained as arising from the proximity of a massive neighbour rather than differences in z_{form} .

The main difference of note is in the velocity anisotropy – if we consider the left-most panels, we see that the latest-forming DORCHA haloes are mildly preferentially radial ($\beta \lesssim 0.2$) at $r \gtrsim 0.5R_{200}$, whereas the earliest-forming ones are more strongly preferentially radial ($\beta \lesssim 0.5$) at $r \simeq R_{200}$.

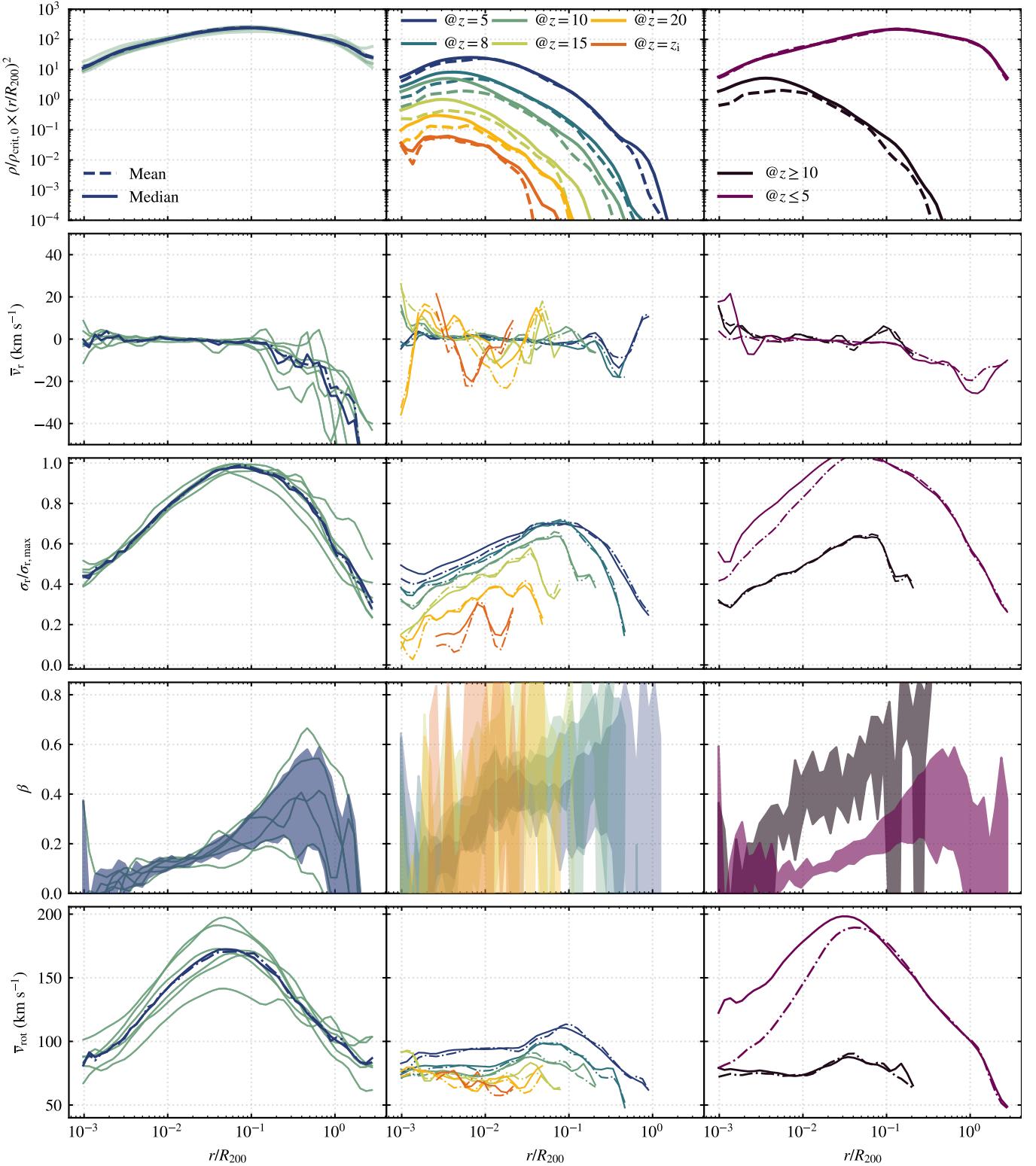


Figure 6. Radial profiles for the early-DORCHAS: Similar to Figure 5 but only for 5 of the earliest-forming DORCHAS ("early" kind). These are DORCHA_02 ($z_{\text{form}} = 2.22$), DORCHA_19A ($z_{\text{form}} = 2.21$), DORCHA_06 ($z_{\text{form}} = 1.93$), DORCHA_10 ($z_{\text{form}} = 1.72$), and DORCHA_05 ($z_{\text{form}} = 1.69$). Note that all of them, except for DORCHA_19A are isolated DORCHAS. Takeaway: The mean and median trends are very similar to the full set of DORCHAS (see Figure 5).

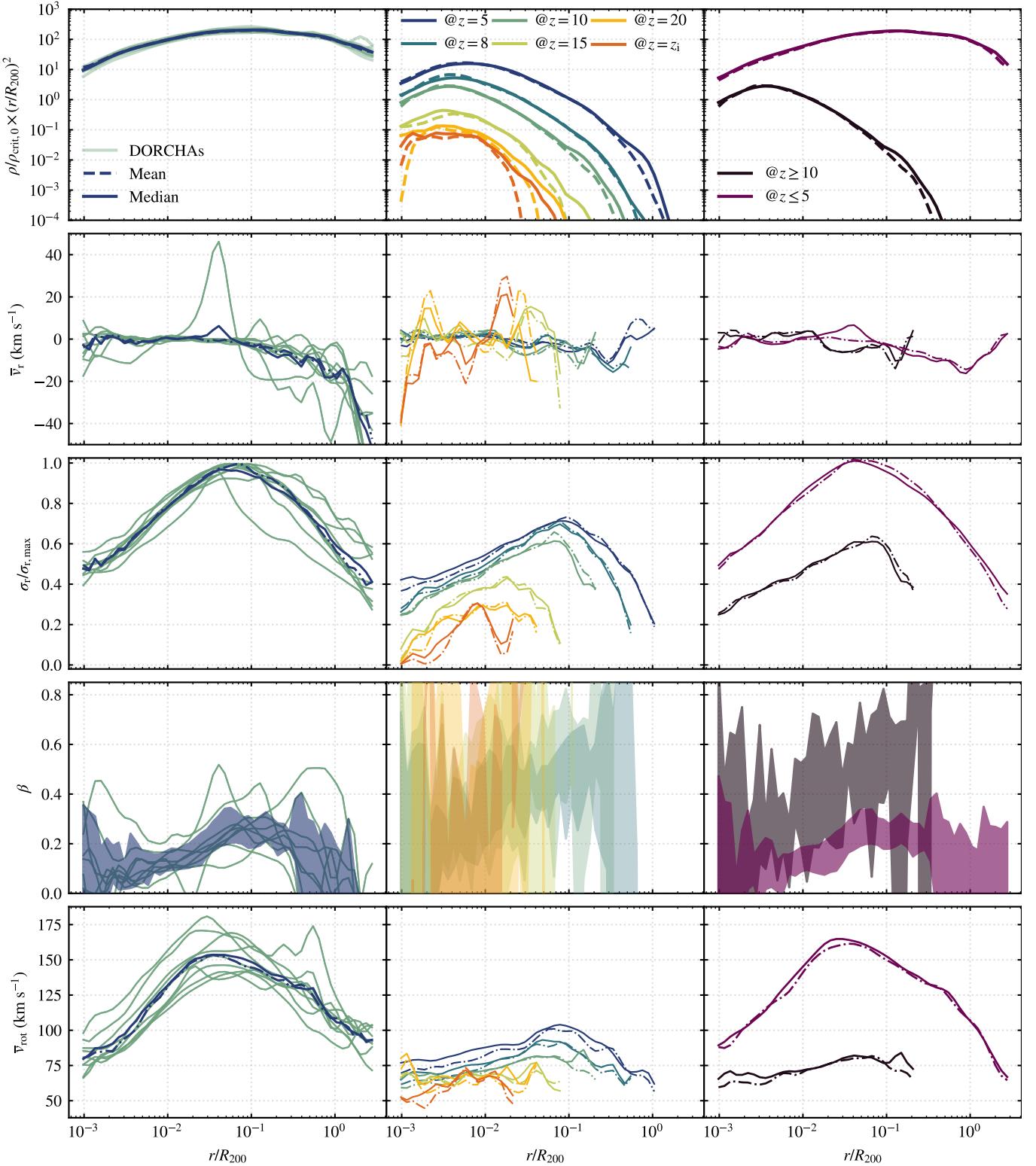


Figure 7. Radial profiles for the late-DORCHAS: Similar to Figures 5 & 6, but only for 5 of the late-forming DORCHAS. These are DORCHA_01 ($z_{\text{form}} = 0.28$), DORCHA_18A ($z_{\text{form}} = 0.63$), DORCHA_19B ($z_{\text{form}} = 0.66$), DORCHA_17A ($z_{\text{form}} = 0.89$), and DORCHA_20A ($z_{\text{form}} = 1.0$), almost all of which are paired DORCHAS. *Takeaway:* The primary difference difference cf. Figure 6 is the $\beta(r)$ profile. Any other minute differences in kinematics can be explained by the presence of a nearby massive neighbour.

However, this reflects environmental differences and the influence of a nearby massive halo.

Note that if we select haloes by z_{10} rather than $z_{\text{form}} \equiv z_{50}$, we find that the results are unchanged – this reflects the relatively quiescent assembly histories of the haloes in our sample, such that rank ordering of the distribution is similar for either z_{10} or z_{50} .

3.2 Phase Space Structure of High Redshift Remnants

The results of the previous section have highlighted that material associated with haloes that collapsed at $z \geq 5$ is

- (i) concentrated in the innermost parts of the halo today, with a density profile that declines as r^{-4} with increasing radius;
- (ii) in dynamical equilibrium (i.e. $\bar{v}_r \simeq 0$);
- (iii) dynamically colder, as measured by σ_r , and has a more strongly preferentially radial velocity distribution, as measured by β , than the bulk of the material that forms the halo.

These results derive from spherically averaged radial profiles, but what kind of structures might we expect in phase space?

In Figures 8 and 9, we show the phase space distribution of halocentric radius, (r) , in h^{-1} kpc against radial velocity, (v_r) , in km s^{-1} for the isolated DORCHA haloes at $z=0$ (plots for the paired haloes can be found in the Appendix A). The colour-coding in Figure 8 indicates the phase-space density at a particular value of (r, v_r) . We include only material that is gravitationally bound to the primary subhalo of the FoF group, and show contours enclosing 90 per cent of the material associated with high redshift, colour-coded by the redshift at which it was identified. In Figure 9, we zoom in on these contours for greater clarity.

Figure 8 reveals that the DORCHA haloes contain a wealth of fine phase space structure. Multiple caustics – broadly symmetric ridges of phase density extending from $-200 \text{ km s}^{-1} \lesssim v_r \lesssim 200 \text{ km s}^{-1}$ and most evident within the inner $0.25 R_{200}$ ($50 h^{-1}$ kpc) – can be found in many of the haloes. For example, the caustics in DORCHA_01 and DORCHA_15 are sharply defined, although in most cases the caustics are less distinct. This fine phase space structure provides an integrated record of the orbits of disrupting substructure within the halo potential during its assembly history (Henriksen & Widrow 1997). At radii of order $0.5 - 1 R_{200}$ ($100 - 200 h^{-1}$ kpc), the phase space structure becomes more asymmetric, and we see ridges with skewed $v_r < 0$ that can be traced to spurs and bifurcations on the haloes’ edges (such as in DORCHA_01 and DORCHA_15). This reflects the interaction of the haloes with their larger scale environment.

Figure 9 allows for closer inspection of the phase space distribution of the material associated with high redshift progenitors, and finds broad consistency across the sample. The contours are symmetric and centred on $v_r = 0 \text{ km s}^{-1}$, as we would expect from the spherically averaged profiles in the previous section. The $z = 5$ contours span the range from $-200 \text{ km s}^{-1} \lesssim v_r \lesssim 200 \text{ km s}^{-1}$ and extend to $\sim 0.1R_{200}$ ($r \simeq 10 - 20 h^{-1}$ kpc), while the oldest material ($z \geq 15$) spans $-100 \text{ km s}^{-1} \lesssim v_r \lesssim 100 \text{ km s}^{-1}$ within $\sim 0.005R_{200}$ ($r \simeq 1 h^{-1}$ kpc). There is some fine-scale structure evident

in the contours – for example, spurs in DORCHA_09 and DORCHA_15 and localised clumps in DORCHA_08, and are indicative of spatially coherent overdensities. However, on the whole, these phase space distributions confirm that the bulk of the material associated with the high redshift progenitors is confined to the innermost parts of the potential; dynamically mixed and relaxed; and relatively dynamically cold.

3.3 Orbital Structure of High Redshift Remnants

The previous results demonstrate that material associated with high redshift remnants is centrally concentrated within the MWAs at $z = 0$ – precisely in the region where the galaxy disc will form. This suggests that the oldest remnant material should be well mixed within the inner regions of the galaxy, and its spatial and kinematic structure is likely to be modified in the flattened disc potential of the galaxy. The size of this effect will correlate with how long this material has been present within the inner regions of the halo.

To assess this, we carry out explicit orbit tracking of all the particles associated with high redshift progenitors using the algorithm of Walker et al. (in prep). Our method follows particles from $z = 0$ to high redshifts across consecutive simulation snapshots and logs when each one experiences a pericentric passage relative to the centre of the MWA’s main progenitor; it does this by identifying a sign change in a particle’s radial velocity (relative to the MWA’s main progenitor’s centre) from negative to positive. To filter out pericentres associated with orbits within substructure, the algorithm selects only radial velocity sign changes that occur after a particle has advanced through a minimum angle, θ_{\min} , around the halo centre since its last sign change. We use $\theta_{\min} = \pi/2$.

Figure 3.3 shows the probability distributions (PDFs) of the minimum pericentric distances of the particles that make up the material associated with high redshift remnants, i.e. those tagged at $z \geq 10$. Here, we track their orbits from $z = 5$ and determine the minimum pericentric distance of a particle during this period, as a fraction of R_{200} at $z = 0$.

We indicate paired and isolated DORCHAS by pink and green curves; the subset of early- and late-forming DORCHAS are indicated by dot-dashed and dashed curves respectively. As we anticipate, the distributions of minimum pericentric distances rise sharply at $r \lesssim 0.01R_{200}$ and peak in the range $\sim 0.001 - 0.01R_{200}$. Combined with Figure 9, which suggest that the typical 90 per cent contour occurs at $\sim 0.1R_{200}$, this indicates that this material will follow high eccentricity orbits. There is no obvious distinction between haloes that are paired or isolated, or early- or late-forming, which is consistent with our previous results.

3.4 Looking for the remnants of high redshift star formation in the Milky Way

The results we have presented have made clear that, regardless of the details of the assembly history of the MWA host halo, the material associated with the remnants of MWA high redshift progenitors will be centrally concentrated in the innermost parts of the potential at $z = 0$; it will be dynamically relaxed; and the oldest components will relatively dynamically cold compared to their surroundings. Almost all of this material associated with high redshift remnants ends up in

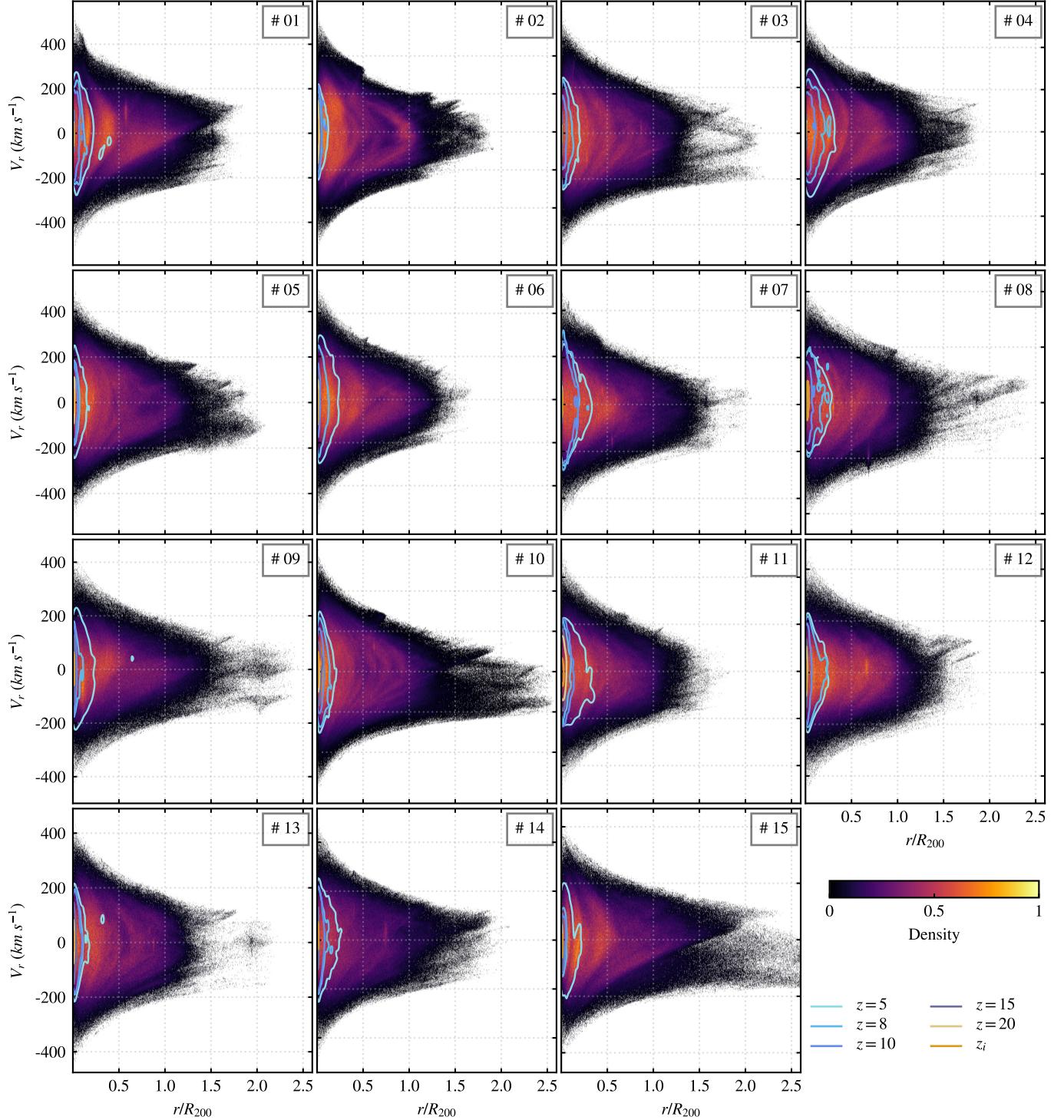


Figure 8. Phase plot for the isolated DORCHAS: Shown are the phase-space distribution, i.e. the radial velocity (v_r) against the radial distance (r), for the entire halo. The lines represent the 90 per cent contour lines of the distribution for particles that were tagged at the indicated redshifts. *Takeaway:* Multiple caustics can be clearly seen in many of the DORCHAS. It is evident that most of the tagged high- z progenitor sites have sunk to the centre of the halo (see also Figure 9).

the main subhalo of our DORCHA haloes. This is consistent with previous simulations (e.g. Tumlinson 2010; Starkenburg et al. 2017; Gao et al. 2010; Horta et al. 2023; but see also El-Badry et al. 2018) and observations (see e.g. Arentsen et al. 2020; Sestito et al. 2022; Ardern-Arentsen et al. 2024).

Our working assumption is that this material traces the phase space of the sites of high redshift star formation. We remind the reader that the correspondence between old stars and metal-poor stars is not one-to-one: metal-poor stars can arise from enriched gas after the first star formation episodes,

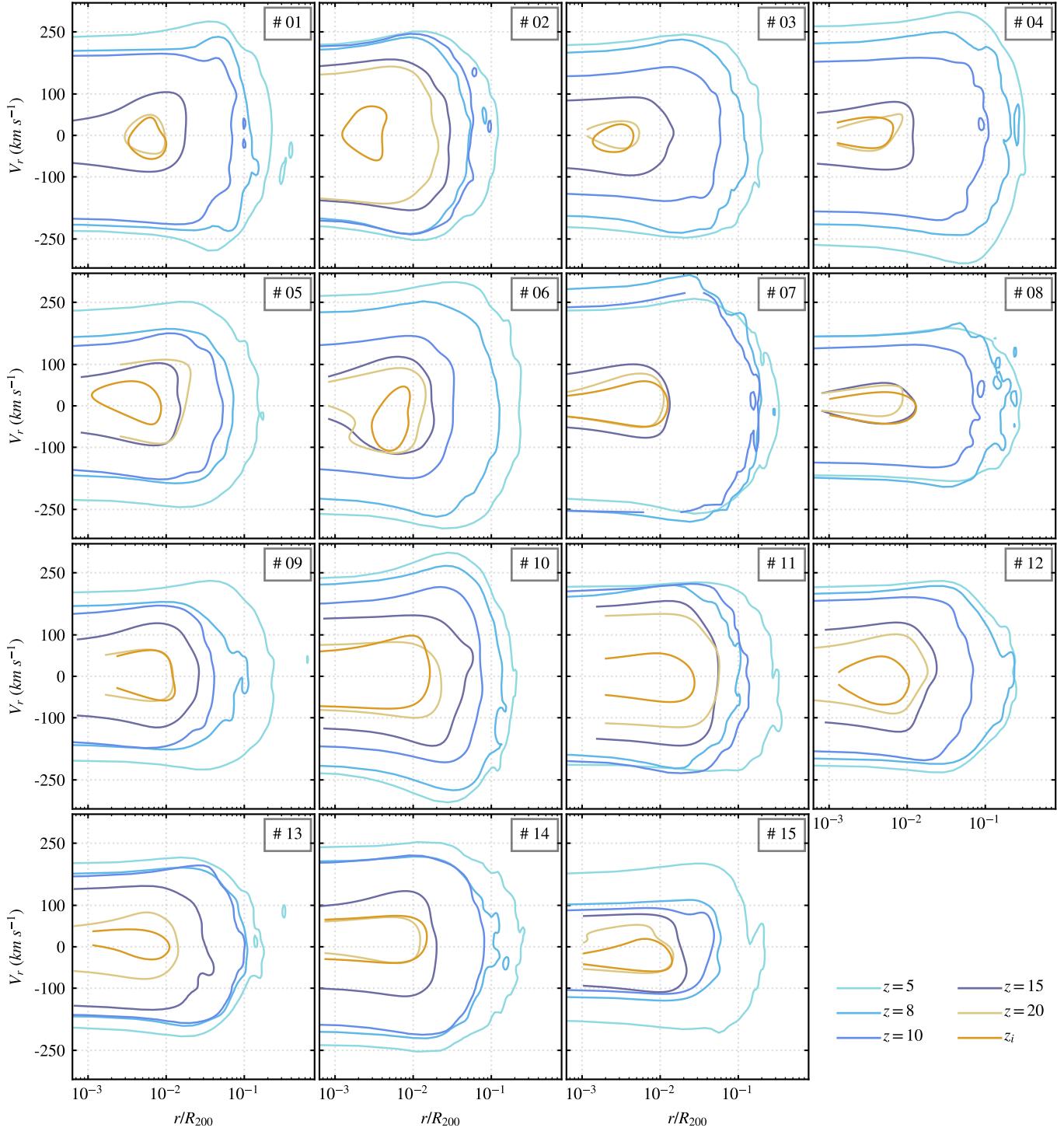


Figure 9. Phase plot for the isolated DORCHAs: A zoom in on the contours shown in Figure 8. *Takeaway:* The bulk of the material associated with the high- z progenitors is confined to the innermost parts of the potential, and is dynamically mixed, relaxed and cold.

and stellar age is more reliably inferred from formation time than from metallicity alone. In Figure 11, we quantify the fraction of material that was in place at $5 \leq z \leq z_i$ within $r = 15 h^{-1}\text{kpc}$ ($\sim 0.1R_{200}$) of the centre of each DORCHA halo and plot the corresponding PDFs based on the 25 haloes in our sample. If we focus on the oldest components ($z \geq 15$),

the PDFs are sharply peaked at 100 per cent, implying that all of these particles are within $r = 15 h^{-1}\text{kpc}$ for almost all the DORCHAs. In contrast, for $5 \leq z \leq 10$, the peak is at 80 and 70 per cents – there is thus a scatter, albeit a small one, amongst the DORCHAs as to where these progenitors end up at $z = 0$. Nevertheless, the trend is for the distributions

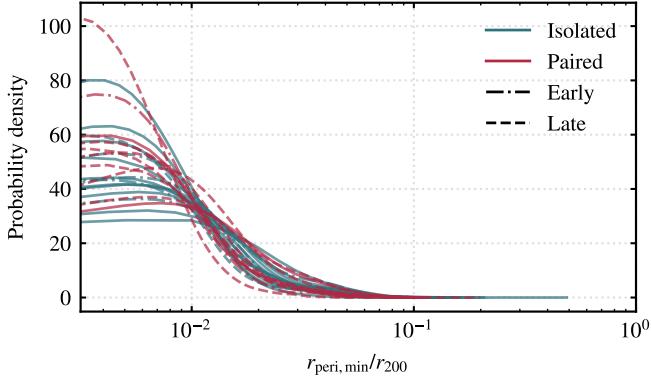


Figure 10. Pericentric distribution: The probability distribution of the minimum pericentric distances of the "old" (i.e. $z \geq 10$) tagged particles for $z \leq 5$. The pink (light green) curves denote paired (isolated) DORCHAS. The dashed (dash-dotted) represent late (early) DORCHAS. *Takeaway:* For all the DORCHAS, most of the particles that we tag at high- z haloes have interacted with the centre of the main subhalo. There is no distinguishable trend that stands out for different categories of DORCHAS. This is likely a consequence of the fact that these particles have sunk into the potential well of their haloes and are thus more robust to external perturbations such as mergers and accretions.

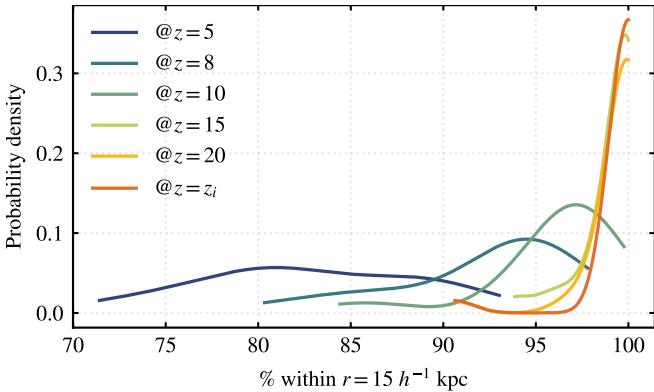


Figure 11. Spatial distribution of tagged particles: The probability distribution of the present-day fraction of tagged particles that are within $r \leq 15 h^{-1} \text{kpc}$ of the halo centre for all 25 of the DORCHA haloes. The different line corresponds to the redshift at which the particles were tagged. *Takeaway:* Most of the tagged particles end up near the halo centre and this is more apparent the higher the redshift of tagging is.

of material associated with older components to be found predominantly within $r = 15 h^{-1} \text{kpc}$ for all the DORCHAS.

This tells us that the oldest stars should be found within the innermost parts of a MWA's halo, within the galaxy disc, most likely contributing to the nuclear bulge. This would be consistent with the kinematics of the $z \geq 10$ components in our DORCHA haloes, which are in line with estimates of the Milky Way's bulge (cf. velocity dispersion of $\sim 100 \text{ km s}^{-1}$, Valenti et al. 2018). It is also suggested by the mass density profile (cf. Figure 5), which drops off as r^{-4} , although the amount of mass contained in these components is likely to be negligible.

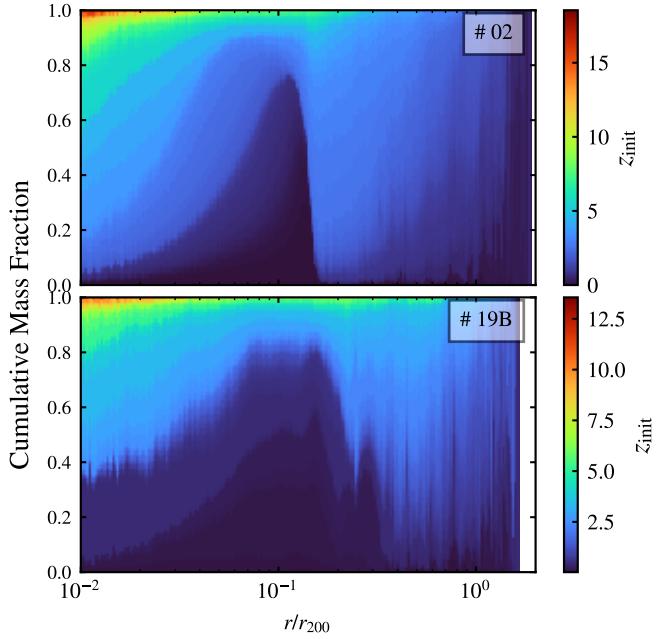


Figure 12. Cumulative mass fraction of the progenitor cores: We show the cumulative mass fraction as a function of the normalized radial distance, r/r_{200} , for DORCHA_02 and DORCHA_19B at $z = 0$, coloured by their formation redshift (z_{init}). At each radius, the colour corresponds to the redshift at which a given fraction (from 0 to 1) of the local cumulative mass formed. *Takeaway:* This reveals how the formation times of particles vary with radius – for instance, most of the mass within inner regions of the halo, say $r/r_{200} \leq 5 \times 10^{-2}$, are dominated by the particles comprising the high- z (i.e. $z \geq 6$) progenitors.

As a further check, in Figure 12, we show the cumulative mass fraction of the progenitors for DORCHA_01 and DORCHA_19B⁵. For this, we tag and track across all 123 snapshots the 10 per cent most-bound particles of all subhaloes of the progenitor FoF group of our DORCHAS. For every particle, we also keep track of the first redshift (z_{init}) at which it forms part of the central core of a progenitor. At $z = 0$, we restrict our attention to the main subhalo, and compute the cumulative mass fraction as a function of both radial distance, r/r_{200} , as well as z_{init} . This allows us to answer: *For a given radius, when did particles that make up a certain fraction of the mass first appear as part of a progenitor?* It is clear from both subplots of Figure 12 that amongst the progenitor material, more than 60 per cent of the mass in the inner regions ($r/r_{200} \leq 5 \times 10^{-2}$) of the halo are contributed by particles that were in the central cores of progenitors at $z \geq 6$.

⁵ Note that DORCHA_01 & DORCHA_19B are, respectively, the earliest-forming isolated and latest-forming paired DORCHAS. This should help us in maximizing any difference in the trend for our analysis here. We confirm that our discussion here applies to all the DORCHAS.

4 CONCLUSIONS

Galactic archaeology – searching for fossil signs of early galaxy formation and evolution in the Local Universe – has the potential to help constrain the first epochs of star formation. By identifying the descendants of the first stars in and around the Galaxy, we can get an understanding of the integrated interaction history of these objects with the Milky Way as well as how and where they assembled in the early Universe. A number of surveys are ongoing and/or are planned to exploit the vast data associated with individual stars. It is important then that appropriate simulations are on-hand to help interpret them. Moreover, simulations can help design better survey parameters – for instance, random blind surveys have a low probability of finding these low-metallicity stars (Ness et al. 2013a), particularly in the crowded and dust-obscured inner parts of the MW (Howes et al. 2016; Lucey et al. 2019; Arentsen et al. 2020). It is therefore vital to identify any spatial, kinematical, and/or dynamical signatures that these old components may exhibit.

Motivated by these considerations, we have introduced the dark-matter-only DORCHA cosmological simulation suite of Milky Way analogues (MWAs). The DORCHA haloes have been chosen to have masses similar to that of the DM halo of the Milky Way. The suite currently consists of 20 simulations – 15 DORCHAS are in isolated environments (DORCHA_01 through DORCHA_15) and the rest are in pairs (DORCHA_16 through DORCHA_20) giving a total of 25 MWA haloes. Though we are agnostic to the particular merger histories of the MWAs while selecting them for resimulation from the parent simulation, DORCHA_00, our isolation criteria results in them having a relatively smooth merger history. Nevertheless, the DORCHAS exhibit a wide range of formation redshifts, from $z_{\text{form}} = 0.28$ to 2.22.

We identify, tag, and track the most-bound 10 per cent particles – which trace the deepest parts of the gravitational potential well – for all high- z progenitor subhaloes of our DORCHA haloes down from redshifts ($z \in [25, 20, 15, 10, 8, 5]$). We then study and explore the spatial and kinematical distributions of these tagged set of particles in the present day $z = 0$ DORCHAS. This include radial profiles (such as the density, mean radial velocity, velocity dispersion, velocity anisotropy, and rotational velocity) as well as their phase space structure. We further sub-categorize the haloes into early- and late-DORCHAS based on their formation redshift z_{form} . This enables us to explore if there is any potential impact of recent mergers and/or accretion on our conclusions – in that a halo that formed most of its mass much earlier than average would relatively have had less disruptive mergers and/or accretion compared to one that formed much later. For $z \leq 5$, we track the orbits and compute the minimum pericentric distance for all the “old”, i.e. $z \geq 10$, tagged particles. This enables us to constrain the amount of interactions that these progenitor sites would have had with the central regions of the main subhalo. Further, we study the spatial location of these progenitor sites near the solar neighbourhood. For this we estimate the fraction of these particles that are within $15 h^{-1}$ kpc of the centre of the halo. We also study the cumulative mass fraction of the DORCHA haloes at $z = 0$ that is composed of the progenitor particles as a function of both radial distance as well as formation redshift.

The main conclusions from our analyses are

- From the radial profiles, we show that the particles that constituted the central regions of the high- z progenitors of our DORCHA haloes are centrally concentrated, mostly within the virial radius $R_{200\text{crit}}$. This effect is more pronounced the higher the redshift of tagging is.

- The radial density profiles of these progenitors fall off as $\sim r^{-4}$ (Hernquist 1990), signifying that these components are centrally concentrated. The median behaviour for these components is $\bar{v}_r \simeq 0 \text{ km s}^{-1}$ which we interpret as them having been the part of the larger halo for at least a few dynamical times.

- We do not find any significant change when we repeat our analyses for the early- and late-DORCHAS – except for a mild variation for the velocity anisotropy profile $\beta(r)$. This is surprising as we had used the formation redshift, z_{form} , as a proxy for the amount of recent accretion/mergers that a halo would have experienced. There are two likely reasons for this: our halo-selection criteria has resulted in relatively quiet growth histories for the DORCHA. Second, even at z_{50} or even z_{10} , a typical MWA is $\sim 10^{10} h^{-1} \text{M}_{\odot}$. The main subhaloes of such systems are robust to external perturbations.

- We see this notably when we compare the results for isolated and paired DORCHAS. Both of these categories exhibit similar trends. In fact, they are identical within the virial radius R_{200} (see also Garrison-Kimmel et al. 2014). This is good news if one is mostly concerned with simulations of MWAs; an expensive simulation of two paired haloes can be avoided in this case.

- The DORCHAS exhibit unique features in their phase space structures; but the overall trends are very similar. There are well-defined caustics, and distinct shells that reflect materials from different accretion and/or merger events.

- Interpreting the radial profiles alongside the phase space structures, we establish that, regardless of the details of the assembly history, the remnants of the high- z progenitors are centrally concentrated in the innermost parts of the potential at $z = 0$, and dynamically relaxed and cold.

We equate the central cores – traced by the 10 per cent most-bound particles – of these gravitationally bound DM structures as sites where gas can cool down efficiently to trigger star formation. Our results suggest that all such star formation sites and the stellar material associated with them would have been accreted by the main subhalo of the MWAs. By $z = 0$ they would have sunk to the centre of the potential. Likely locations for metal-free PopulationIII stars and their remnants as well as low-metallicity stars that would have formed from their supernova ejecta are thus the central cores of the MWAs. We stress, however, that the correspondence between old stars and metal-poor stars is not one-to-one: metal-poor stars can arise from enriched gas after the first star formation episodes.

As we noted in § 1, the DORCHA suite provides the foundation for further study with the SOLAS suite of hydrodynamical galaxy formation simulations. We expect that the inclusion of baryons, and the impact of radiative cooling and explosive forms of feedback, will modify the gravitational potential, as well as influencing which progenitors might host star formation and the degree to which they might remain as coherent substructure in phase space. However, we argue that the DORCHA suite is sufficient to provide robust insights into key structural and kinematic properties of the old stellar

population of the Milky Way and its relationship to galaxy formation in the early Universe. We will revisit this in forthcoming papers.

ACKNOWLEDGEMENTS

This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project #CE170100013. This work was also undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS-enabled capability supported by the Australian Government. SB acknowledges the Dr Albert Shimmins Fund for a Postgraduate Writing-Up Award, and support from grant PID2022-138855NB-C32 funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU, and project PPIT2024-31833, cofunded by EU–Ministerio de Hacienda y Función Pública–Fondos Europeos–Junta de Andalucía–Consejería de Universidad, Investigación e Innovación.

Software citations: We thank the MUSIC, AREPO and GADGET-4 development teams for making these codes public. This research also relies heavily, and with great thanks, on the PYTHON (Van Rossum & Drake Jr 1995) open source community, in particular, NUMPY (Harris et al. 2020), MATPLOTLIB (Hunter 2007), SCIPY (Virtanen et al. 2020), H5PY, JUPYTER (Granger & Pérez 2021), and PANDAS (pandas development team 2023). The visualizations in Figures 1 & 2 were generated with YT (Turk et al. 2011).

DATA AVAILABILITY

All simulation data, including snapshots, group and subhalo catalogues, and merger trees will be made publicly available in the near future at solas-sims.github.io/data_products. In the meantime, the data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Ardern-Arentsen A., et al., 2024, *Monthly Notices of the Royal Astronomical Society*, 530, 3391
- Arentsen A., et al., 2020, *Monthly Notices of the Royal Astronomical Society: Letters*, 491, L11
- Barbuy B., Chiappini C., Gerhard O., 2018, *Annual Review of Astronomy and Astrophysics*, 56, 223
- Barkana R., Loeb A., 2001, *Physics Report*, 349, 125
- Barnes J., Hut P., 1986, *Nature*, 324, 446
- Bland-Hawthorn J., Gerhard O., 2016, The galaxy in context: Structural, kinematic, and integrated properties, doi:10.1146/annurev-astro-081915-023441, www.annualreviews.org
- Boylan-Kolchin M., 2023, *Nature Astronomy*, 7, 731
- Buch D., Nadler E. O., Wechsler R. H., Mao Y.-Y., 2024, *The Astrophysical Journal*, 971, 79
- Bullock J. S., Johnston K. V., 2005, *The Astrophysical Journal*, Volume 635, Issue 2, pp. 931-949., 635, 931
- Callingham T. M., et al., 2019, *Monthly Notices of the Royal Astronomical Society*, 484, 5453
- Carollo D., et al., 2010, *The Astrophysical Journal*, Volume 712, Issue 1, pp. 692-727 (2010)., 712, 692
- Casey C. M., et al., 2024, *The Astrophysical Journal*, 965, 98
- Chandro-Gomez A., et al., 2025, *Monthly Notices of the Royal Astronomical Society*, 539, 776
- De Jong R. S., et al., 2019, Published in *The Messenger* vol. 175, pp. 3-11, 9 pages
- De Silva G. M., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 2604
- Diemand J., Madau P., Moore B., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 367
- El-Badry K., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 480, 652
- Frebel A., Norris J. E., 2015, *Annual Review of Astronomy and Astrophysics*, 53, 631
- Freeman K., Bland-Hawthorn J., 2002, *Annual Review of Astronomy and Astrophysics*, 40, 487
- Furlanetto S. R., Peng Oh S., Briggs F. H., 2006, *Physics Reports*, 433, 181
- Gaia Collaboration et al., 2016, *Astronomy & Astrophysics*, 595, A1
- Gao L., Theuns T., Frenk C. S., Jenkins A., Helly J. C., Navarro J., Springel V., White S. D. M., 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 1283
- Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, *Monthly Notices of the Royal Astronomical Society*, 438, 2578
- Granger B. E., Pérez F., 2021, *Computing in Science & Engineering*, 23, 7
- Griffen B. F., Dooley G. A., Ji A. P., O’Shea B. W., Gómez F. A., Frebel A., 2018, *Monthly Notices of the Royal Astronomical Society*, 474, 443
- Hahn O., Abel T., 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 2101
- Harris C. R., et al., 2020, *Nature*, 585, 357
- Helmi A., 2020, *Annual Review of Astronomy and Astrophysics*, vol. 58, p.205-256, 58, 205
- Henriksen R. N., Widrow L. M., 1997, *Physical Review Letters*, 78, 3426
- Hernquist L., 1990, *The Astrophysical Journal*, 356, 359
- Horta D., et al., 2023, *Monthly Notices of the Royal Astronomical Society*, 527, 9810
- Howes L. M., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 4241
- Howes L. M., et al., 2016, *Monthly Notices of the Royal Astronomical Society*, 460, 884
- Hunter J. D., 2007, *Computing in Science & Engineering*, 9, 90
- Ivezic Z., et al., 2019, *The Astrophysical Journal*, 873, 111
- Knebe A., Gill S. P. D., Kawata D., Gibson B. K., 2005, *Monthly Notices of the Royal Astronomical Society: Letters*, 357, L35
- Koch A., et al., 2008, *The Astrophysical Journal*, 689, 958
- Lucey M., et al., 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 2283
- Ludlow A. D., Schaye J., Bower R., 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 3663
- Madau P., Kuhlen M., Diemand J., Moore B., Zemp M., Potter D., Stadel J., 2008, *The Astrophysical Journal*, 689, L41
- Majewski S. R., et al., 2017, *The Astronomical Journal*, Volume 154, Issue 3, article id. 94, 46 pp. (2017)., 154, 94
- McConnachie A., Murowinski R., Salmon D., Simons D., Côté P., 2014, *Ground-based and Airborne Telescopes* V, 9145, 914519
- Mellema G., et al., 2013, *Experimental Astronomy* 2013 36:1, 36, 235
- Morales M. F., Wyithe J. S. B., 2010, *Annual Review of Astronomy and Astrophysics*, 48, 127
- Navarro J. F., et al., 2010, *Monthly Notices of the Royal Astronomical Society*, 402, 21
- Nebrin O., Giri S. K., Mellema G., 2023, *Monthly Notices of the Royal Astronomical Society*, 524, 2290
- Ness M., et al., 2013a, *Monthly Notices of the Royal Astronomical*

- Society, 430, 836
 Ness M., et al., 2013b, *Monthly Notices of the Royal Astronomical Society*, 432, 2092
 Peñarrubia J., Gómez F. A., Besla G., Erkal D., Ma Y.-Z., 2015, *Monthly Notices of the Royal Astronomical Society: Letters*, 456, L54
 Planck Collaboration et al., 2020, *Astronomy & Astrophysics*, 641, A6
 Pritzl B. J., Venn K. A., Irwin M., 2005, *The Astronomical Journal*, 130, 2140
 Sestito F., et al., 2022, *Monthly Notices of the Royal Astronomical Society*, 518, 4557
 Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 1105
 Springel V., 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 791
 Springel V., et al., 2008, *Monthly Notices of the Royal Astronomical Society*, 391, 1685
 Springel V., Pakmor R., Zier O., Reinecke M., 2021, *Monthly Notices of the Royal Astronomical Society*, 506, 2871
 Stark D. P., Topping M. W., Endsley R., Tang M., 2025, Observations of the First Galaxies in the Era of JWST, doi:10.48550/arXiv.2501.17078, <http://arxiv.org/abs/2501.17078>
 Starkenburg E., Oman K. A., Navarro J. F., Crain R. A., Fattahi A., Frenk C. S., Sawala T., Schaye J., 2017, *Monthly Notices of the Royal Astronomical Society*, 465, 2212
 Thomas G. F., Battaglia G., Grand R. J. J., Aguiar Álvarez A., 2025, arXiv e-prints, p. arXiv:2504.10398
 Tolstoy E., Hill V., Tosi M., 2009, *Annual Review of Astronomy & Astrophysics*, vol. 47, Issue 1, pp.371-425, 47, 371
 Tumlinson J., 2010, *The Astrophysical Journal*, 708, 1398
 Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, *The Astrophysical Journal Supplement Series*, 192, 9
 Valenti E., et al., 2018, *Astronomy & Astrophysics*, 616, A83
 Van Rossum G., Drake Jr F. L., 1995, Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam
 Virtanen P., et al., 2020, *Nature Methods*, 17, 261
 Weinberger R., Springel V., Pakmor R., 2020, *The Astrophysical Journal Supplement Series*, 248, 32
 pandas development team T., 2023, pandas-dev/pandas: Pandas, doi:10.5281/zenodo.7857418, <https://zenodo.org/record/7857418>

APPENDIX A: PHASE-SPACE STRUCTURE OF THE PAIRED DORCHAS

In Figures A1 and A2 we show the phase-space structure of the paired DORCHAS (DORCHA_16 to DORCHA_20). Figure A1 is similar to figure 8 where we show the full phase-space information of all the particles that comprise the corresponding DORCHA halo. Similar to the isolated DORCHAS (see 3.2) almost all of the particles that we tag at high- z end up concentrated towards the centre of the halo. This is brought out clearly in figure A2 where, just like in figure 9, we focus on the central portions of the halo and show the 90 percentile contour lines of the distributions of the tagged particles.

This paper has been typeset from a TeX/LaTeX file prepared by the author.

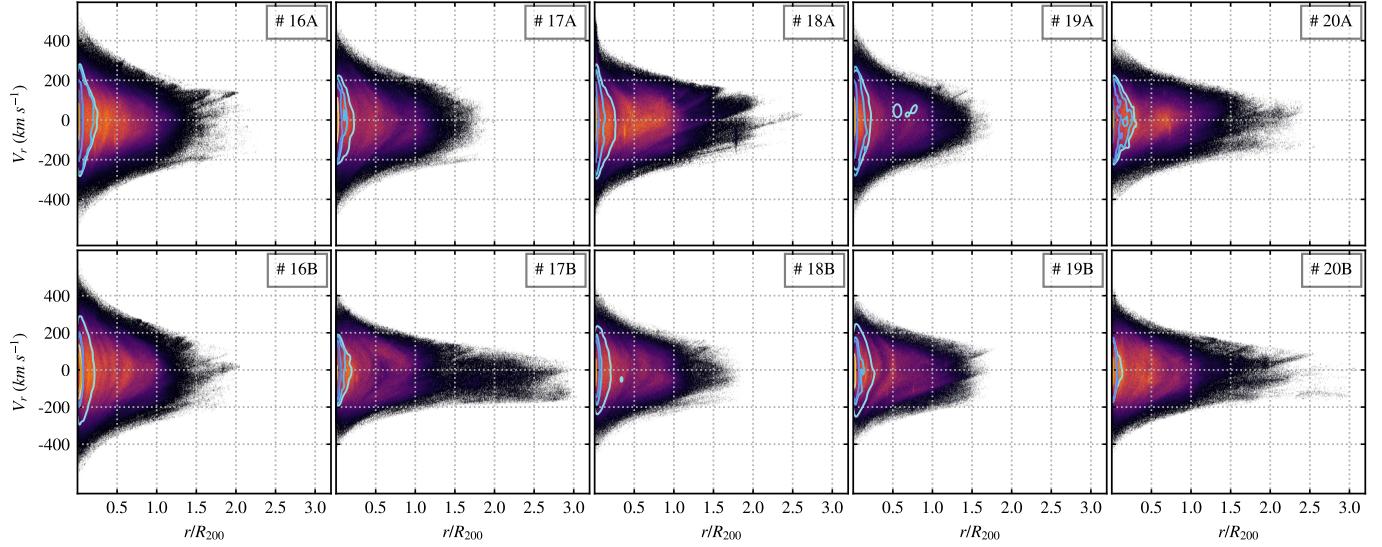


Figure A1. Phase plot for the paired DORCHAS:: Same as the Figure 8 but for the paired DORCHA simulations (DORCHA_15-20). The more massive of the pair (at $z = 0$) is labelled A is shown in the top-row. *Takeaway:* We see a very similar trend for the paired DORCHAS relative to the isolated ones. Here as well, it is clear that most of the tagged particles are concentrated towards the halo centre.

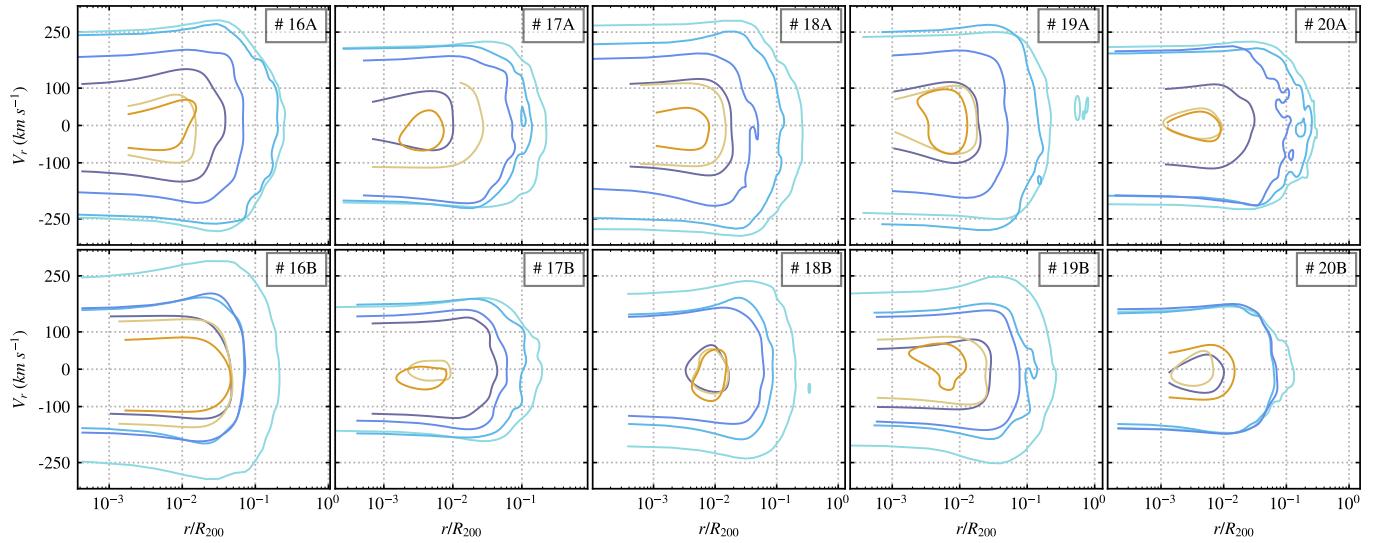


Figure A2. Phase plot for the isolated DORCHAS:: Shown are the phase-space distribution of entire halo, i.e. the radial velocity (v_r) against the radial distance ((r)). The contours represent the 90 per cent contour lines that encompass the phase-space distribution of the tagged particles. *Takeaway:* Figure shows the variety of phase-space distributions of the DORCHA haloes. It is clear that most of the tagged particles end up within $\sim 100h^{-1}\text{kpc}$ of the halo centre which is much lower than the typical R_{200} .