Astronomy & Astrophysics

The survey of planetary nebulae in Andromeda (M31)

VII. Predictions of a major merger simulation model compared with chemodynamical data of the disc and inner halo substructures

C. Tsakonas[1],⋆ , M. Arnaboldi[2] , S. Bhattacharya[3,4] , F. Hammer[5], Y. Yang[5], O. Gerhard[6] ,
R. F. G. Wyse[7] , and D. Hatzidimitriou[1]

[1] Department of Astrophysics, Astronomy & Mechanics, Faculty of Physics, National and Kapodistrian University of Athens, Panepistimiopolis, Zografou 15784, Greece
[2] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
[3] Inter University Centre for Astronomy and Astrophysics, Ganeshkhind, Post Bag 4, Pune 411007, India
[4] Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK
[5] GEPI, Observatoire de Paris, PSL Research University, CNRS, Place Jules Janssen, F-92190 Meudon, France
[6] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstraße, 85748 Garching, Germany
[7] Department of Physics & Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA

ABSTRACT

Context. The nearest giant spiral, the Andromeda galaxy (M31), exhibits a kinematically hot stellar disc, a global star formation episode ∼2–4 Gyr ago, and conspicuous substructures in its stellar halo that are suggestive of a recent accretion event.
Aims. Recent chemodynamical measurements in the M31 disc and inner halo can be used as additional constraints for N-body hydrodynamical simulations that successfully reproduce the disc age-velocity dispersion relation and star formation history as well as the morphology of the inner halo substructures.
Methods. We combined an available N-body hydrodynamical simulation of a major merger (mass ratio 1:4) with a well-motivated chemical model to predict abundance distributions and gradients in the merger remnant at $z = 0$. We computed the projected phase space and the [M/H] distributions for the substructures in the M31 inner halo, namely, the Giant Stellar Stream (GSS) and the North-East (NE) and Western (W) shelves. We compared the chemodynamical properties of the simulated M31 remnant with recent measurements for the M31 stars in the inner halo substructures.
Results. This major merger model predicts (i) multiple distinct components within each of the substructures; (ii) a high mean metallicity and large spread in the GSS and NE and W shelves, which explain various photometric and spectroscopic metallicity measurements; (iii) simulated phase space diagrams that qualitatively reproduce various features identified in the projected phase space of the substructures in published data from the Dark Energy Spectroscopic Instrument (DESI); (iv) a large distance spread in the GSS, as suggested by previous tip of the red giant branch measurements; and (v) phase space ridges caused by several wraps of the secondary as well as up-scattered main M31 disc stars that also have plausible counterparts in the observed phase spaces.
Conclusions. These results provide further strong and independent arguments for a major satellite merger in M31 ∼3 Gyr ago and a coherent explanation for many of the observational results that make M31 appear so different from the Milky Way.

Key words. galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: interactions – galaxies: kinematics and dynamics – Local Group

1. Introduction

As the closest large galaxy, the Andromeda galaxy (M31) was the object of many early groundbreaking studies in extragalactic astronomy (e.g. Hubble 1929; Rubin et al. 1970). Due to its large total mass (∼1.5×10^{12} M$_\odot$; see Bhattacharya 2023 and references therein) and proximity – 773 kpc from the Milky Way, MW; Conn et al. 2016 – it is an ideal laboratory for galactic archaeology studies through wide-field photometric identification (e.g. Ibata et al. 2001; Ferguson et al. 2002; Dalcanton et al. 2012; Williams et al. 2017) and resolved spectroscopy (e.g. Guhathakurta et al. 2006; Escala et al. 2020, 2022; Dey et al. 2023b) of red giant branch (RGB) stars.

The most comprehensive imaging survey of M31 is the Pan-Andromeda Archaeological Survey (PAndAS; McConnachie et al. 2018). It covers a ∼400 sq. degree area and targets the RGB stars within the virial radius of M31. PAndAS revealed an intricate pattern of stellar substructures that populate the inner halo of M31 and whose origin is readily associated with accretion event(s).

The most striking among the substructures is the Giant Stellar Stream (GSS), a stellar structure extending to more than ∼80 kpc (Ibata et al. 2001, 2004; McConnachie et al. 2003, 2018; Cohen et al. 2018) in projected distance from the centre of M31. It is connected to the second brightest substructure in the M31 inner halo, the North-East (NE) shelf, via stellar streams (Merrett et al. 2003). The GSS is the most prominent stellar stream in the Local Group and a clear sign of galactic

interaction. It has thus become apparent that, in stark contrast with the MW, which had a quiet accretion history (e.g. Wyse et al. 2001; Pillepich et al. 2014; Helmi 2020) over the last ~10 Gyr, M31 seems to have undergone significant accretion events in its recent past (McConnachie et al. 2009).

Following the discovery of the GSS, several simulation attempts were carried out to reproduce its morphology as the result of a minor merger event (i.e. when the mass ratio between the two progenitor galaxies is less than 1:10, see Ibata et al. 2004; Geehan et al. 2006; Font et al. 2006; Fardal et al. 2006, 2007, 2008, 2012, 2013; Mori & Rich 2008; Sadoun et al. 2014; Kirihara et al. 2017; Milošević et al. 2022, 2024). While differences exist among the various minor merger models, they characteristically consider that a satellite with a total mass of ~10^9 M$_\odot$ plunged into the gravitational potential of M31 following a radial orbit. Many of these models have been successful at reproducing the morphology of the primary tidal features in the inner halo of the galaxy; namely, the GSS, the NE shelf (Ferguson et al. 2002; Ibata et al. 2005), and the Western (W) shelf (Fardal et al. 2007). In these models, the GSS appears as a relatively short-lived tidal stream, with the accretion event occurring ~1 Gyr ago. Additionally, they produce wedge-like features in the diagram of the projected distance versus line-of-sight (LOS) velocity (hereafter R_{proj} versus V_{LOS} diagram) for the satellite stars in these substructures (Fardal et al. 2007; Kirihara et al. 2017; Milošević et al. 2024). Such features in the projected phase space are typical of a radial merger event (Merrifield & Kuijken 1998; Hendel & Johnston 2015). Indeed, coherent kinematic structures (streams, chevrons, wedges[1]) were later observed in the R_{proj} versus V_{LOS} diagram of stars in the NE shelf (Escala et al. 2022) and the rest of the inner halo substructures (Dey et al. 2023b, hereafter D23), providing additional supporting evidence for a recent merger event. However, such minor merger events typically do not influence the kinematics of the disc of the more massive galaxy (Martig et al. 2014).

The disc of M31 displays interesting properties that are difficult to explain under the minor merger assumption. Dorman et al. (2015) found that the age-velocity dispersion relation for the stars in the disc of M31 is considerably steeper than that reported (e.g. Nordström et al. 2004) for the MW's disc. The average LOS velocity dispersion at radii larger than $R = 10$ kpc is an increasing function of stellar age, rising from 30 km s^{-1} for the main sequence stars up to 90 km s^{-1} for ~4 Gyr old RGB stars. A subsequent study from Bhattacharya et al. (2019a) using planetary nebulae (PNe[2]) surveyed in M31 (Bhattacharya et al. 2019b) found an increase in the rotational velocity dispersion within the old stellar population, from $\sigma_\phi \sim 61$ km s^{-1} for PN progenitor stars with ages less than 2.5 Gyr to $\sigma_\phi \sim 101$ km s^{-1} for PN progenitors older than 4.5 Gyr, which are associated with the thin disc and the thicker disc, respectively. The velocity lag displayed by the two populations independently supports the age ordering (Bhattacharya et al. 2019a). The large amount of kinetic energy input needed to heat the disc to such dispersion requires a merger with a massive companion (see Hopkins et al. 2009).

Bhattacharya et al. (2022) measured PNe oxygen and argon abundances as a function of the disc radius (out to 30 kpc). Inter-

preting the PN abundances with tailored galactic chemical evolution models in the oxygen-to-argon ratio versus argon abundance plane, Arnaboldi et al. (2022) showed that the M31 disc experienced an infall of metal-poor gas ~2–4 Gyr ago that diluted the interstellar medium (ISM), out of which the subsequently younger stars were formed (see also Kobayashi et al. 2023 for galactic chemical enrichment models pertaining to this event). Thus, the last significant accretion event in M31 was a star-forming merger in which gas from the satellite and/or the outer M31 gas disc was brought inwards. These results align with the widespread burst of star formation ~2 Gyr ago identified with deep Hubble Space Telescope (HST) observations of the resolved stellar population in the M31 disc (Williams et al. 2017) by the Panchromatic Hubble Andromeda Treasury (PHAT) survey (Dalcanton et al. 2012) and in pencil beam observations of the outer regions of the disc of M31 (Bernard et al. 2012, 2015).

A kinematically hot thick disc such as that in M31 can be produced in a wet major merger event (Brook et al. 2004; Hopkins et al. 2009). The hot thick M31 disc and the burst of star formation ~2–3 Gyr ago together with the intricate morphological characteristics of M31's inner halo were reproduced within the context of a single recent wet galaxy merger event by Hammer et al. (2018, hereafter H18). They presented N-body hydrodynamical simulations of a major satellite merger with a mass ratio of 1:4 (hereafter major merger for brevity) with significant gas content. The merger resulted from the collision between two disc galaxies, and the coalescence of the nuclei occurred in the time interval 1.8–3 Gyr ago. The parameters of the simulations were optimised to reproduce M31's inner disc and central (bulge, bar) properties as well as the substructures close to the disc and within the inner halo (see Sect. 2.1 for further details). The H18 major merger simulations also predict that the G1-clump substructure observed at the southern edge of M31's disc (Ferguson et al. 2002) was formed from a substantial perturbation of the outer M31 disc. The kinematics of the PNe co-spatial with the G1-clump were later found to be consistent with that predicted in the major merger scenario (Bhattacharya et al. 2023). Such a substructure formed out of perturbed host disc material is yet to be observed in minor merger simulations.

While some of the morphological and kinematic features of the R_{proj} versus V_{LOS} diagram of the GSS and NE and W shelves can be reproduced by minor merger scenarios (e.g. Fardal et al. 2006; Font et al. 2006), they fail to reproduce the two kinematically cold components (KCCs) observed in several inner fields of the GSS (Kalirai et al. 2006; Gilbert et al. 2007, 2009) as well as the spread in LOS distance within the stream (Conn et al. 2016). H18 simulations predict a broad LOS velocity distribution and several kinematically distinct components in the GSS region (Bhattacharya et al. 2023) that are associated with debris generated during the pericentre passages of the satellite (see Sect. 4.1 for further details).

Recent observations of RGB stars in the inner halo of M31, including the GSS and NE and W shelves, acquired from the Dark Energy Spectroscopic Instrument (DESI) multi-object spectrograph (D23) provide for the first time stellar kinematics and metallicity measurements with a uniform wide-field coverage over tens of square degrees. Combined with metallicity estimates from recent multi-object spectroscopic observations from pencil beam surveys in various regions (Escala et al. 2019, 2020, 2021, 2022; Gilbert et al. 2019, 2020) and additional photometric surveys in the inner halo (Bernard et al. 2015; Tanaka et al. 2010; Conn et al. 2016; Ogami et al. 2025), these datasets provide additional chemodynamical constraints against which the

---

[1] Wedges and chevrons refer to triangular-shaped patterns in the R_{proj} versus V_{LOS} diagram of stars in the substructures. Wedges refer to filled triangles and chevrons to empty triangles.

[2] PNe are emission line nebulae in a short-lived late stage of stellar evolution for stars with initial masses ~0.8–8 M$_\odot$ (see the review by Kwitter & Henry 2022 and references therein). The PNe thus span a wide range of parent stellar population ages (~100 Myr to ~10 Gyr).

major merger simulations of H18 can be tested. Our investigation aims to extend the comparison to the simulated phase space metallicity distributions of the inner halo substructures with the recent measurements in M31.

In the present work, we build on a viable major merger model from H18 and extend the previous analysis. As there are new observations of the age and abundance distributions for the stellar tracers in the M31 disc and substructures, the major merger model can be tested further. We do so by setting a well-motivated chemical distribution for the two progenitor galaxies at the beginning of the simulation. We make use of the recent PNe sample measurements in M31's disc from Bhattacharya et al. (2022) and observational constraints from the stellar mass-metallicity relation (hereafter MZR; see Maiolino & Mannucci 2019 and references therein) for disc galaxies of similar mass as M31 and observed at redshifts comparable with that of the simulated merger. We then follow the disc particle forwards through the merger and compare the chemodynamical properties of the remnant disc and the three prominent inner halo substructures (GSS and NE and W shelves) of M31's simulated analogue with extensive observational data.

The paper is organised as follows. In Sect. 2, we describe the main properties of the simulation selected from H18. Sect. 3 presents the steps undertaken to constrain the initial metallicity distribution in the two progenitor galaxies. The stellar substructures in the inner halo of the simulated galaxy have multiple components that we analyze separately in Sect. 4. In Sect. 5, we compare the $R_{\rm proj}$ versus $V_{\rm LOS}$ diagram of the GSS and NE and W shelves of our simulated merger remnant with the recent observational results for M31. In Sect. 6, we compare the predicted metallicity values with photometric and spectroscopic metallicity measurements in the inner halo substructures. We state our main conclusions in Sect. 7.

2. A major merger simulation for the evolution of M31

On the basis of the results obtained from numerical simulations that showed thin discs to be rebuilt after gas-rich galaxy mergers (Hammer et al. 2005, 2009; Hopkins et al. 2009), H18 utilised ∼300 high-resolution simulations to investigate and reproduce the morphological properties of M31. These simulations consisted of two bulgeless disc galaxies, both with substantial gas discs, in addition to the stellar discs. H18 have let free most of the parameters within a large range (see their Table 1), such as the mass ratio from 1:5 to 1:2. However, H18 did not consider minor merger cases (including mass ratio significantly larger than 1:10), conversely to the works of Fardal et al. (2013) and Kirihara et al. (2017) for instance, where the mass of the secondary galaxy is assumed to be ∼5 × 10⁹ M⊙. The rationale for the adopted strategy is illustrated in H18 and in the introduction of the current work. To recap, only a major merger can reproduce the observed properties of M31's disc: a strong star forming episode 2–3.5 Gyr ago (see Bernard et al. 2012, 2015; Williams et al. 2017), the steep age-velocity dispersion relation (Dorman et al. 2015), and the 10 kpc ring (Lewis et al. 2015). The mass merger ratio (constrained from 1:4 to 1:5) is independently found in the investigation carried out by Bhattacharya et al. (2019a), who used the relation between the disc scale height (H_z) and the satellite-to-disc-mass ratio ($M_{\rm sat}/M_{\rm disc}$) described by Hopkins et al. (2008), to explain the dynamically hot 4.5 Gyr-old population in M31's disc (with an $H_{z\,4.5\rm Gyr} \simeq 0.86$ kpc; Dalcanton et al. 2015).

In addition to the observed properties of the M31 disc, there are compelling observational constraints on the timing of the merger event in M31 and its first pericentre passage occurring ∼7–8 Gyr ago, coming from the clustering of the quenching time of dwarf spheroidal galaxies ($10^5\,{\rm M_\odot} < {\rm M_*} < 10^7\,{\rm M_\odot}$) in M31; see D'Souza & Bell (2021), Savino et al. (2025).

We choose model #336, one of the best M31 analogues and the highest resolution simulation of H18, with 20 million particles, as it did the best job of matching the observations at the time it was published. The mass resolution is 13 750 M⊙ and 55 000 M⊙ per baryon and dark matter particle respectively. By utilizing the major merger model, we aim to make predictions for the chemodynamical phase space structure of the resulting M31 analogue, with particular emphasis on the prominent substructures in the inner stellar halo, including the GSS (Sect. 4.1).

The simulation is deployed for a total time length of 7.5 Gyr. Each progenitor is initially modelled by two components, a dark matter halo (comprising 80% of the total mass) and a thin disc that includes both stars and gas. The scale length of the gas discs is three times that of the stellar discs and the scale height of all discs (stellar and gaseous) is equal to 1/10 of the scale length of the stellar discs (i.e. the gas/stellar discs of each progenitor are assumed to be thin and have the same scale height). Both discs have flat rotation curves, which reach $V_{\rm max} \simeq 220$ km s⁻¹ for the main and $V_{\rm max} \simeq 140$ km s⁻¹ for the secondary disc.

2.1. General properties and time evolution of the M31 simulations

H18 used the N-body hydrodynamical code GIZMO to simulate the collision between the two bulgeless, thin discs with a mass ratio of 1:4. The primary features of the dynamical evolution of the merger are as follows. The first pericentric passage occurs at a lookback time of ∼7 Gyr. Due to their high relative velocity, the two discs reside far from each other for the next ∼3.5 Gyr, until the second pericentric passage occurs. After the second pericentric passage, the two galaxies do not drift far apart from each other. The third pericentric passage occurs ∼1 Gyr later, followed by a full-fledged merging process. Multiple subsequent pericentric passages at short time intervals follow and the satellite progenitor is completely disrupted. The coalescence of the nuclei takes place 1.8–3 Gyr ago, and from a lookback time of ∼2 Gyr, the two galaxies are fully merged into a single galaxy.

Following the second pericentric passage ∼3.5 Gyr ago, the subsequent pericentric passages of the secondary release tidally distorted material (stellar debris from both progenitors) in the inner halo of the remnant. This tidal debris shows up as overdensities along the plane of the infalling satellite's orbit, resulting in overlapping loops that are destined to merge with M31. The simulation predicts that the superposition of these loops along the LOS and their projections on the sky produce the observed properties of the inner halo substructures in velocity and position.

During the collision, the ensuing gravitational torques within the gas brought by the satellite and the gas from the main progenitor resulted in the ignition of star formation in the main progenitor disc ∼2–3 Gyr ago, as was observed by Williams et al. (2017) and Bernard et al. (2012, 2015). During the final stages of the merger, many short pericentric passages of the secondary heat the main progenitor's pre-existing stellar distribution, resulting in the formation of the kinematically hot thick disc in the remnant. The thick disc is mostly comprised of stars older than the collision event (ages >4 Gyr). A residual core from the
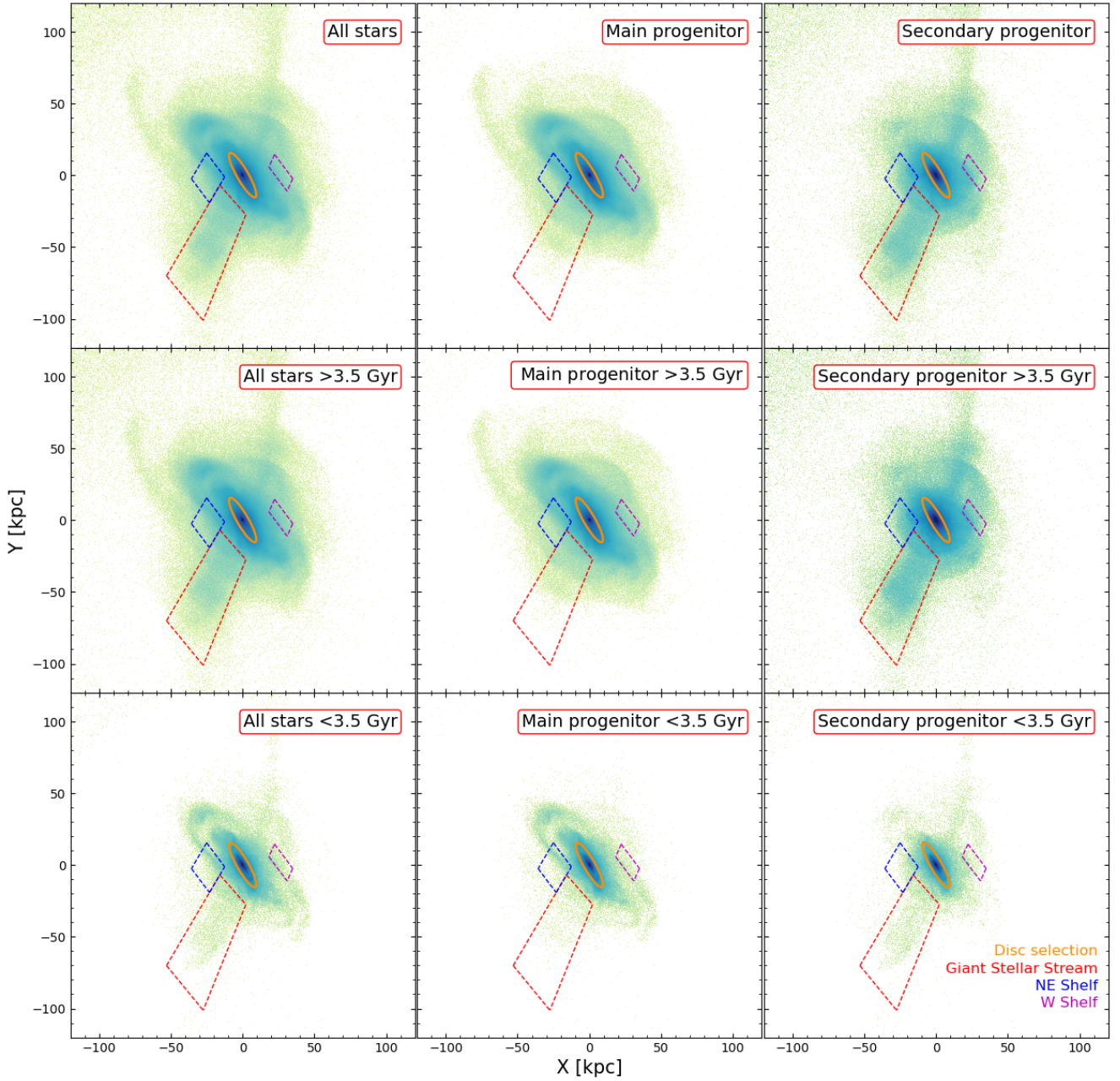
Fig. 1. Spatial density map of the star particles at the end of the merger simulation. The galaxy is rotated and projected onto the sky plane according to the inclination and PA of M31. The selected areas for the M31 disc, the GSS and NE and W shelves are indicated with different colours. Upper row: All simulated stellar particles [left], main progenitor stellar particles [middle], and secondary progenitor stellar particles [right] are shown in independent panels. Middle row: Same differentiation, but only for stars older than 3.5 Gyr. Lower row: Same differentiation, but only for stars younger than 3.5 Gyr.

secondary progenitor is absent in the remnant disc, as the central parts of the secondary progenitor completely dissolve during the numerous short pericentric passages, and its – initially less bound – stellar outskirts are deposited into inner halo substructures (GSS and NE and W shelves).

Overall, the major (1:4) mass merger model for the evolution of M31 successfully predicts the steep age-velocity dispersion relation (Dorman et al. 2015; Bhattacharya et al. 2019a), the star formation burst in the disc (Williams et al. 2017), the chemodynamics of the G1 clump (Bhattacharya et al. 2023) and the clustering of the quenching time of the dwarf spheroidal galaxies ~8 Gyr ago (D'Souza & Bell 2021),

associated with the first pericentre passage of the merging satellite.

2.2. Morphology of the merger remnant

The distribution of the stars in the remnant at the completion of the simulation is portrayed in Figure 1. Star particles are demarcated separately for each progenitor (all particles [left], main progenitor particles [middle], and secondary progenitor particles [right]). Middle and lower rows further differentiate stars into two age groups, with 3.5 Gyr being the defining age; this specific choice is discussed within the context of the star formation

history (SFH) of the disc of M31 in Appendix B. Given the timing of the merger (the second pericentric passage occurred ~3.5 Gyr ago), stars older than 3.5 Gyr formed before the strong interactions, while younger stars formed from gas mainly in the rebuilt thin disc (see Sect. 2.1). From the beginning of the simulation, stellar and gaseous particles of each progenitor are labelled as main or secondary galaxy particles and retain this label throughout the simulation (even if a gaseous particle is converted into a star within the remnant disc).

Plotted as a 2D histogram, Figure 1 encapsulates the predicted stellar distribution at the end of the simulation ($z = 0$). The orange ellipse shows the boundary of the remnant galaxy disc. We introduce this selection based on the estimated length of the major (30 kpc; Bhattacharya et al. 2019a) and minor (3.99 kpc for an inclination of 77°; Geehan et al. 2006) axes of the galaxy. For the appropriate rotation of the modelled disc, we assume a position angle (PA) of 30° (instead of the commonly cited PA = 38°), which properly reproduces the observed morphology of the disc. We note that the thin stellar disc in the simulation has a scale length of 3.6 kpc (H18). Yin et al. (2009) estimate a disc scale length of 6.08 kpc for M31. We use this measurement as a baseline, and we scale our ellipsoidal selection for M31's disc accordingly (scaling down the major and minor axes in the model accordingly). The selection of the inner halo substructures (GSS and NE and W shelves) is based on their projected sky positions relative to the centre of M31, which reproduce the locations of their observed counterparts (McConnachie et al. 2018); our demarcations are almost identical to the spatial selections made for the GSS and NE and W shelves in Bhattacharya et al. (2023). For an extensive discussion on other prominent features evident in Figure 1 (like the stream at the north-west part of the remnant, dubbed the Counter GS), we refer the reader to H18 and Bhattacharya et al. (2023).

In Figure 1 [upper, middle], stellar particles from the more massive progenitor populate mainly the central parts of the galaxy, including the inner and outer parts of the disc, and the inner halo. In Figure 1 [upper, right], particles from the secondary galaxy form the tidal streams and shell-like substructures in the halo of the remnant galaxy, with the GSS being mostly formed from stellar debris of the secondary galaxy. Separate illustrations of the two age groups are given in the middle and lower rows. These distributions illustrate that the bulk of the inner halo is composed of old stars. In the lower panel, the distribution of young stars in the model traces the recent episode of star formation confined to the inner disc.

3. Predicting oxygen abundances and metallicity gradients in the M31 disc

The H18 simulations compute the enrichment of gas particles due to star formation, in addition to computing the orbits of both star and gas particles in the time-dependent gravitational potential. Implementing a suitable set of assumptions, one can then compute and compare the stellar metallicity (expressed in terms of [M/H] or [Fe/H] for α/Fe = 0.0) and oxygen abundance in various regions of the disc, substructures, and inner halo of the simulated remnant with the abundance measurements of the oxygen/iron content from stellar and gaseous probes.

The metallicity of each gas and stellar particle at the beginning of the simulation is a free parameter. The N-body hydrodynamical models compute i) the metallicity of the star particles formed out of the gas distribution during the simulations and ii) the enrichment of the gas particles due to star formation. For those stars associated with the initial stellar discs of the progen-

itors, there is no further enrichment. These stars maintain the value assigned ab initio throughout the entire simulation. Star particles formed from gas that is initially in the secondary progenitor maintain their label (i.e. secondary particle), even if the star formation episode occurred within the remnant disc.

3.1. The enrichment process of the interstellar medium in the H18 simulations

The approach to compute the chemical enrichment process of the ISM in the simulation is described in detail in Cox et al. (2006). It assumes a yield of 0.02 per solar mass of stars formed. The enriched material is instantly recycled and depends only on the instantaneous star formation rate of each gas particle. The star formation rate is determined by the local gas density and is calculated individually for each gas particle, with metal enrichment carried out within such particles. The metal enrichment is smoothed over the smoothing kernel of the corresponding gas particles. Specifically for model 336 (H18) used in this paper, the smoothing kernel of star-forming particles has a median kernel size of 0.16 kpc (ranging from 0.05 to 0.18 kpc). Metal diffusion is not explicitly modelled as a separate process, but estimates of the properties of gas particles (enrichment, among others) are smoothed over the smoothing kernel, effectively mimicking the effects of metal diffusion (Okamoto et al. 2005; Tornatore et al. 2007; Wiersma et al. 2009; Roca-Fàbrega et al. 2021). Each gas particle carries its metal content, and when new star particles form, they inherit the metallicity from their parent gas particles at that point in time. Once a star is born, its metal content remains unchanged.

This particular enrichment recipe is akin to the enrichment carried out through core-collapse supernovae (SN II); no delayed enrichment, analogous to supernovae type Ia (SN Ia), is implemented in the code.

We convert, when needed, metallicity to oxygen abundance given the solar oxygen abundance ($12+\log(\text{O/H})_\odot = 8.69$; Asplund et al. 2009) and, assuming that the solar elemental abundance is universal (Salaris & Cassisi 2005), through the following equation:

$$[\text{M/H}] = 12 + \log(\text{O/H}) - 8.69. \tag{1}$$

We also express the stellar and gas metallicity in terms of iron abundance: [M/H] = [Fe/H], assuming α/Fe = 0.0 (solar alpha enrichment).

3.2. Constraints on the metallicity distribution of the initial stellar and gaseous discs

Each galaxy, representing the main and secondary progenitor, consists of two distinct structures, that is a stellar and a gaseous disc, which have different disc scale lengths. Thus, the initial oxygen abundance (or, equivalently, [M/H]) distribution must be set for each of the four discs at the beginning of the simulation. We adopt a linear relationship between radius and oxygen abundance values, with a negative gradient. Hence, the general functional form of the initial oxygen distribution in any of the discs at the beginning of the simulation is

$$(12 + \log(\text{O/H}))_{\text{ini},R} = \nabla (12 + \log(\text{O/H})) \times R + (12 + \log(\text{O/H}))_{R=0}, \tag{2}$$

where $\nabla (12+\log(\text{O/H}))$ is the adopted (negative) gradient, R is the distance of each particle (stellar or gaseous) from the galaxy

Table 1. Initial conditions for the simulation of H18 for the main and the secondary progenitor.

Parameter	Main	Secondary
$M_\star$ [M$_\odot$]	6.63×10^{10}	0.66×10^{10}
$\log(M_\star/\mathrm{M}_\odot)$	10.82	9.82
h_s [kpc]	2.8	2.8
h_g [kpc]	8.4	8.4
f_gas	0.5	0.8
$(12+\log(\mathrm{O/H}))_{<1R_e}$ [dex]	8.706 ± 0.15	8.518 ± 0.15
$\nabla(\log(\mathrm{O/H}))$ [dex/kpc]	-0.1 ± 0.05	-0.1 ± 0.05

Notes. $M_\star$ is the stellar mass of each galaxy. h_s is the scale length of the stellar discs, h_g is the scale length of the gas discs, and f_gas is the fraction of gas mass to the total baryonic mass in each progenitor (the baryonic mass is 20% of the total mass). $(12+\log(\mathrm{O/H}))_{<1R_e}$ are the values from the MZR measured for galaxies at $z \simeq 1$ (Rodrigues et al. 2008) and for a stellar mass equal to the mass of each progenitor respectively. The uncertainties reported are calculated as the $1 \times \sigma$ dispersion within the relevant redshift bin. The spread in the oxygen gradient values is derived from Figure 9 of Curti et al. (2020a).

centre in the plane of the disc, and $(12+\log(\mathrm{O/H}))_{R=0}$ is the oxygen abundance at $R=0$. Hence, we need to specify i) the gradient $\nabla(12+\log(\mathrm{O/H}))$ and ii) the central $(12+\log(\mathrm{O/H}))_{R=0}$ values for each of the four discs.

To this end, we used the MZR (Maiolino & Mannucci 2019) and the reported oxygen abundance from a large sample of PNe (Bhattacharya et al. 2019b, 2021) in the M31 disc (Bhattacharya et al. 2022) to constrain the initial metallicity distribution for the stellar and gas discs in the simulation. As an additional consistency check, the abundance gradients in the simulated M31 disc at $z = 0$ are compared with the measured abundance gradients determined for different generations of stars (Gregersen et al. 2015; Saglia et al. 2018; Gajda et al. 2021; Bhattacharya et al. 2022).

3.2.1. Constraints from the stellar mass-metallicity relation and observed and simulated metallicity gradients

The MZR links the average of oxygen abundance, within the effective radius (R_e), of the galaxy's gas-phase metallicity and its stellar mass (see Tremonti et al. 2004; Maiolino & Mannucci 2019; Curti et al. 2020b). The simulation by H18 starts at a look-back time of 7.5 Gyr ($z \sim 1$). Since the effective radii and stellar masses of the two progenitors are known quantities in the simulation, we calibrate the average oxygen abundance within the R_e, that is $(12+\log(\mathrm{O/H}))_{<1R_e}$, based on the oxygen abundance of similar stellar mass discs at $z \simeq 1$ using the MZR.

In Table 1, we list the masses, scale lengths, and gas fractions for the two disc progenitors in the simulation. We note that the MZR depends on the adopted method to infer the gas-phase oxygen abundance. The direct (Te) abundance determination using auroral lines, which favour lower $(12+\log(\mathrm{O/H}))_{<1R_e}$ values, is considered the most accurate (see review by Maiolino & Mannucci 2019 and references therein). However, no Te-based MZR is yet available at $z \sim 1$. To account for this observational incongruity, we utilise the study from Rodrigues et al. (2008), where, although they use a strong-line calibration to infer the oxygen abundance of their sampled galaxies, they attain relatively low $(12+\log(\mathrm{O/H}))_{<1R_e}$ values due to their better estimate of the extinction and the underlying Balmer

absorption. The $(12+\log(\mathrm{O/H}))_{<1R_e}$ values for the given progenitors' stellar masses based on this study are also noted in Table 1.

Concerning setting the initial metallicity gradient $\nabla(12+\log(\mathrm{O/H}))$ of the two progenitor galaxies, Curti et al. (2020a) find a significantly wide spread in the measured values for the metallicity gradient with redshift (see their Figure 8). In our current study, we then proceed by adopting the same $\nabla(12+\log(\mathrm{O/H}))$ value equal to -0.1 ± 0.05 for the four discs (see the parameters adopted in Table 2). The arguments in support of this choice are presented in detail in Appendix A.

The $(12+\log(\mathrm{O/H}))_{R=0}$ abundance values are set independently of the gradient values using the constraints set from the $(12+\log(\mathrm{O/H}))_{<1R_e}$ MZR values at $z = 1$ (see Table 2) within the errors. Therefore, the $(12+\log(\mathrm{O/H}))_{R=0}$ values are different for the main and secondary. The metallicity of each star/gas particle is then assigned at the beginning of the simulation according to Equation 2, with a value for the gradient in the range -0.1 ± 0.05, see Table 1. The abundance distributions at $z = 0$ for the different stellar populations in the M31 remnant are then compared with the oxygen abundance distributions for PNe in the M31 disc (see Sect. 3.2.2).

3.2.2. Constraints from oxygen abundance distribution of M31 planetary nebulae

In what follows, we verify that our assumptions for the initial metallicity gradient make predictions which are consistent with measurements of the oxygen abundances from stellar population tracers of different ages in the M31 disc. The survey of PNe in M31, including the disc region from $R_\mathrm{GC} = 3$–30 kpc (Bhattacharya et al. 2019b,a, 2022), provided age estimates and oxygen abundance distributions for stellar progenitors within the M31 disc. PNe with parent stellar populations older than $\sim$4.5 Gyr are associated with the kinematically hotter, thicker disc, while those that are $\sim$2.5 Gyr old are associated with the kinematically colder, thin disc (Bhattacharya et al. 2019a). Additionally, these two PN samples also had distinct oxygen abundance distributions (Bhattacharya et al. 2022).

We thus compared the oxygen abundance distribution of these two age samples with that of their model counterparts. Based on the SFH of the modelled particles versus that determined for the observed M31 disc (Williams et al. 2017), we link the older and younger PN population with that of modelled particles having ages >3.5 Gyr and <3.5 Gyr, respectively. The milestone of 3.5 Gyr for the observed PNe which then separates between thin and thick disc PNe was set independently of the merger timescales in the simulations. The selection of this reference milestone, in conjunction with the disc SFH, is discussed in detail in Appendix B.

We carry our tests whereby the central value is fixed as in Table 1, while we adjust the gradient from the shallowest (-0.05 dex/kpc) to the steepest possible values (-0.15 dex/kpc). The resulting (at $z = 0$) $\langle 12+\log(\mathrm{O/H})\rangle$ of the older stars[3] with ages >3.5 Gyr in the simulation is compared to the measured value for the older PNe. We find a good agreement for $\nabla(12+\log(\mathrm{O/H})) = -0.1$, while for the steeper/shallower values of the gradients, the average $\langle 12+\log(\mathrm{O/H})\rangle$ values of the abundance distributions of the older stars are off by more than $1 \times \sigma$. The comparison of the predicted oxygen abundance distributions

---

[3] Stars in the simulated remnant M31 disc were selected within an elliptical region (as described in Sect. 2.2). A central 3 kpc aperture is masked in the spatial selection of the disc region since the PNe survey does not sample regions of high surface brightness in the centre of M31.

Table 2. Adopted and derived parameters of the major merger model and observational constraints.

Adopted and derived parameters of the model at $z = 1$		
Parameter	Main	Secondary
$(12+\log(O/H))_{R=0}$ [dex]	9.0	8.75
$\nabla(12+\log(O/H))$ [dex/kpc]	−0.1	−0.1
$\langle 12+\log(O/H)\rangle_{Stellar,<1R_e}$ [dex]	8.73	8.48
$\langle 12+\log(O/H)\rangle_{Gas,<1R_e}$ [dex]	8.70	8.45
Derived parameters of the model at $z = 0$		
Parameter	Old stars	Young stars
$\langle 12+\log(O/H)\rangle$ [dex]	8.47	8.67
$\sigma(12+\log(O/H))$ [dex]	0.35	0.36
$(12+\log(O/H))_{R=0}$ [dex]	8.52	8.77
$\nabla(12+\log(O/H))$ [dex/R_d]	−0.06	−0.075
PNe survey (Bhattacharya et al. 2022)		
Parameter	Old PNe	Young PNe
$\langle 12+\log(O/H)\rangle$ [dex]	8.48 ± 0.02	8.57 ± 0.03
$\langle 12+\log(Ar/H)\rangle$ [dex]	6.25 ± 0.02	6.32 ± 0.03
$\sigma(12+\log(O/H))$ [dex]	0.21	0.28
$\sigma(12+\log(Ar/H))$ [dex]	0.20	0.29
$(12+\log(O/H))_{R=0}$ [dex]	8.31 ± 0.05	8.6 ± 0.08
$(12+\log(Ar/H))_{R=0}$ [dex]	6.3 ± 0.05	6.51 ± 0.08
$\nabla(12+\log(O/H))$ [dex/r_d]	0.036 ± 0.018	−0.079 ± 0.036
$\nabla(12+\log(Ar/H))$ [dex/r_d]	−0.03 ± 0.018	−0.109 ± 0.036
RGB survey (Gregersen et al. 2015)		
Parameter	4 Gyr old RGBs	
∇[M/H] [dex/kpc]	−0.02 ± 0.004	

Notes. $(12+\log(O/H))_{R=0}$ and $\nabla(12+\log(O/H))$ are the adopted values for the linear functional form (Eq. (2)) of the oxygen abundance in the initial discs, that is the central value of the oxygen abundance and its gradient along the galactocentric radius. $\langle 12+\log(O/H)\rangle_{Stellar,<1R_e}$ and $\langle 12+\log(O/H)\rangle_{Gas,<1R_e}$ are the oxygen abundance values we compute within one R_e for the initial discs at $z = 1$. The derived parameters for the remnant disc at $z = 0$ are the average oxygen abundance ($\langle 12+\log(O/H)\rangle$) for the two families of modelled stars, older and younger than 3.5 Gyr and its standard deviation (σ), their central abundance (($12+\log(O/H))_{R=0}$), and the oxygen abundance gradient ($\nabla(12+\log(O/H))$). The same parameters for oxygen and argon abundances are reported for the two families (old: >4.5 Gyr; young: ~2.5 Gyr) of PNe from Bhattacharya et al. (2022). The metallicity gradient reported in Gregersen et al. (2015) is also noted.
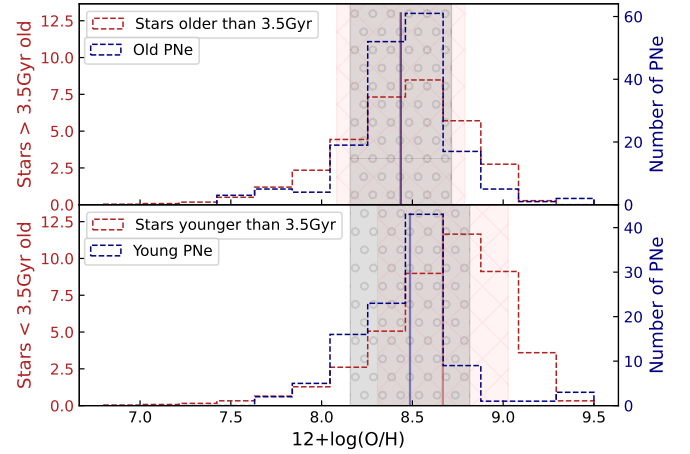

Fig. 2. Upper panel: Comparison of the resulting oxygen abundance of old stars (>3.5 Gyr) in the modelled disc with the oxygen abundance of old PNe from Bhattacharya et al. (2022). Each vertical line specifies the mean values of the two datasets. Shaded regions represent the standard deviation (1×σ) of each dataset. The y-axis [left] is the mass percentage of stars older than 3.5 Gyr in the model. The PNe are plotted according to their number counts in the survey (y-axis [right]). Lower panel: Same as the upper panel but reporting the comparison between younger stars (<3.5 Gyr) in the model and young PNe.

of the simulated stars with those of old and young PNe for a $\nabla(12+\log(O/H)) = -0.1$ gradient is shown in Figure 2.

Figure 2 [upper panel] shows that the oxygen abundance distribution for the older stars[4] is in very good agreement with that of the old PNe. We recall that the cumulative stellar mass of stars older than 3.5 Gyr account for ~85% of the total stellar mass at $z = 0$. Figure 2 [lower panel] shows the corresponding quantities for the younger stars that contribute 15% of the total stellar mass at $z = 0$. The oxygen abundance distribution of the simulated young stars is broader and skewed towards higher oxygen abundance values than that measured for PNe with young stellar progenitors. Quantitatively, the $\langle 12+\log(O/H)\rangle$ of the young simulated stars is about ~0.1 dex more metal-rich than that measured for the young PNe. Such an offset is comparable to the standard deviation values (Table 2). In summary, our approach reproduces the oxygen abundance distribution for the majority of stars (85%) and within 1σ for the remaining 15%.

At this stage, we achieved a good agreement between the simulated old stars (ages >3.5 Gyr) and the oxygen abundance

distribution of the PN with the older progenitors. This is an important result, as the substructures in the inner halo (GSS and NE and W shelves) comprise mainly old stellar populations (Brown et al. 2006, 2007), and we would test the model predictions for their projected phase space and metallicity distributions in Sects. 5 and 6.

For the simulated younger stars, their oxygen distribution turns out to be slightly (~0.1 dex) more metal-rich than that measured for young PNe. Given the limitations of the chemical enrichment treatment as implemented in the H18 simulations, we consider the current setup to provide a satisfactory approximation to within 1σ for the chemical properties of the younger stars in the disc.

3.3. Simulated radial abundance gradients for the different stellar populations in the M31 disc

The chemical evolution of discs of spiral galaxies is reflected in their radial abundance gradient (see Maiolino & Mannucci 2019, and references therein). In this study, we implemented a procedure to predict the radial oxygen gradients for stellar populations for the two different ages from the H18 merger simulations for the M31 remnant disc. Using Equation (1), we convert the oxygen abundances to stellar metallicity gradients, and we utilise different observational data sets as possible benchmarks for comparison.

In Saglia et al. (2018), the authors derived the metallicity gradient from IFU observations of the M31 central regions, within $R_{GC} \sim 4$ kpc, and measured a [M/H] gradient of 0.0 ± 0.03 dex/kpc. The mean [M/H] for their disc population is 0.03 ± 0.03 dex, without postulating an age distinction for the stellar populations under study. In the model, within $R_{GC} \sim 4$ kpc, we report a mean [M/H] of −0.16 dex for all stars, a mean [M/H] of 0.03 dex for stars younger than 3.5 Gyr, and a mean [M/H] of −0.2 dex for stars older than 3.5 Gyr.

Constrained by the aforementioned IFU observations, made-to-measure modelling of M31 by Blaña Díaz et al. (2018) was

---

[4] The oxygen abundance and metallicity ([M/H]) will be used interchangeably to probe the stellar metallicity as well using Eq. (1).
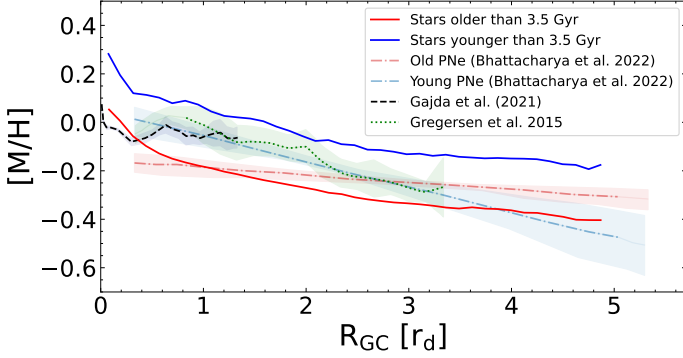
Fig. 3. Comparison of the resulting metallicity along the galactocentric radius for old (red solid curve) and young (blue solid curve) stars in the model with corresponding observations, expressed in units of disc scale length (r_d). For the old and young PNe sample from Bhattacharya et al. (2022), the 12+log(Ar/H) gradient are shown.

utilised by Gajda et al. (2021) to determine the [M/H] radial profile out to ~10 kpc (Figure 3). Gregersen et al. (2015) measured a metallicity gradient (see Table 2) for $R_{GC} \simeq 4$–20 kpc in the M31 disc using stars from the PHAT survey (also, Figure 3). They assumed solar [α/Fe] value and a constant age of 4 Gyr for the RGB stars. Bhattacharya et al. (2022) measured oxygen and argon radial abundance gradients for the two age groups of PN progenitors. Within the radial range of 3–30 kpc, their reported oxygen abundance gradients are shown in Table 2 and plotted in Figure 3, separately for their two PNe groups.

Figure 3 also presents the mean metallicity in the modelled stars for our two distinct age groups as a function of galactocentric radius. The fitted parameters for the radial oxygen abundance (akin to [M/H]) distribution of these two age groups are noted in Table 2. For the young modelled stars, the gradient is consistent with the negative gradient from the young PNe, but with a higher $(12+\log(O/H))_{R=0}$ value. For the older modelled stars, we find a negative gradient, different from the positive gradient found for older PNe, but with a higher $(12+\log(O/H))_{R=0}$ value that leads to a consistent $\langle 12+\log(O/H)\rangle$. As shown in Figure 3, the [M/H] as a function of the galactocentric radius for the two groups of modelled stars brackets the observed distributions from stellar probes of different ages.

As a general remark, we also point out that old stars in the simulation model show a flatter gradient relative to younger stars, as expected in galaxies that experienced mergers (e.g. Zinchenko et al. 2015; Tissera et al. 2019).

3.4. Radial redistribution of gas particles during the major merger event

The metallicity distributions of stellar and gaseous particles in the remnant M31 disc at the end of the merger simulation are the results of the i) initial assignment of the oxygen abundance gradients in the stellar and gaseous discs of both main and secondary progenitors, and contributions from either ii) the triggered star formation or iii) the radial redistribution of the stellar and gaseous particles during the merger event.

To assess the interplay between the adopted initial abundance gradients and the enrichment due to the star formation in the simulation, we estimate the radial redistribution of young stars in the disc from the initial radial distance of the gas particles in the progenitors' discs. In Figure 4, we show the ΔR (i.e., $R_{initial} - R_{final}$) of stars younger than 3.5 Gyr separated based on their initial pro-
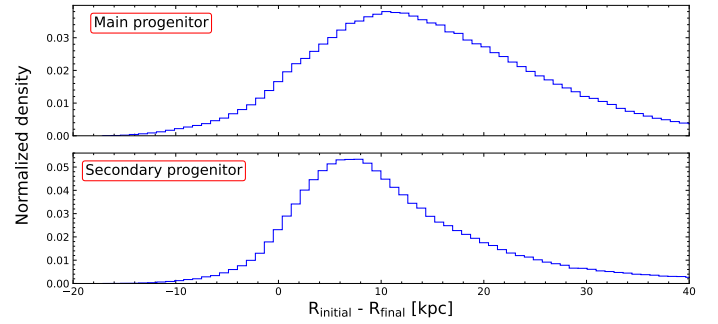
Fig. 4. Normalized density of the distribution of the difference between the initial ($R_{initial}$) and final (R_{final}) galactocentric distances of young (<3.5 Gyr) stars in the two progenitor galaxies. Here, $R_{initial}$ is the distance of each gas particle from the centre of its progenitor (main or secondary) at the beginning of the simulation (7.5 Gyr ago), whereas R_{final} is the galactocentric distance of the stellar particle that spawned from its gaseous progenitor and is situated in the remnant disc. Upper panel: Stars younger than 3.5 Gyr from the main progenitor. Lower panel: Stars younger than 3.5 Gyr from the secondary progenitor.

genitor's association (we remind the reader that star/gas particles retain their initial progenitor identity regardless whether they are converted into stellar particles within the remnant disc).

A preliminary assessment of the spatial rearrangement of stars within the final remnant is as follows: As seen in Figure 4, 93% of young stars (<3.5 Gyr old), from both secondary and main progenitors, have moved inwards ($\Delta R > 0$) from their initial galactocentric radius. We note again, that since we deal with stars younger than 3.5 Gyr, these particles were initially gaseous and were converted into stars during the last 3.5 Gyr of the simulation. Hence, the majority of young stars in the model came initially from the gas particles residing in the outskirts of the progenitors. The initially negative (-0.1 dex/kpc) metallicity gradient led to their lower metal content.

The in-depth assessment of the impact of different mechanisms (e.g. the stellar bar, the energy injection from the secondary, the spiral arms, etc.) on the observed radial redistribution of disc stars caused by the major merger event, will be the topic of a future study.

4. New predictions of the multidimensional phase space structure of the $Z = 0$ remnant in the simulation

In a minor accretion event scenario, the substructures in the M31 inner halo (i.e. GSS and NE and W shelves), are single leading (NE shelf) and trailing (GSS, W shelf) orbital wraps of tidally disrupted stars (e.g. Fardal et al. 2013; Kirihara et al. 2014, 2017) which were formed in the last 1 Gyr. Hence, the shell patterns for the NE and W shelves are predicted to correspond to successive orbital passages of GSS-related tidal debris (Escala et al. 2022). The phase space distributions (projected radius R_{proj} versus LOS velocity V_{LOS}) of their stellar components are therefore smooth (see, for example, Fig. 10 in Escala et al. 2022). Over-densities or ridges may be the "smoking gun" for the possible presence of an intact core of the secondary progenitor, at the location of the NE shelf (see again Fig. 10 in Escala et al. 2022 featuring the simulated projected phase space diagram for the NE shelf in case of the survival of the secondary progenitor's intact core).
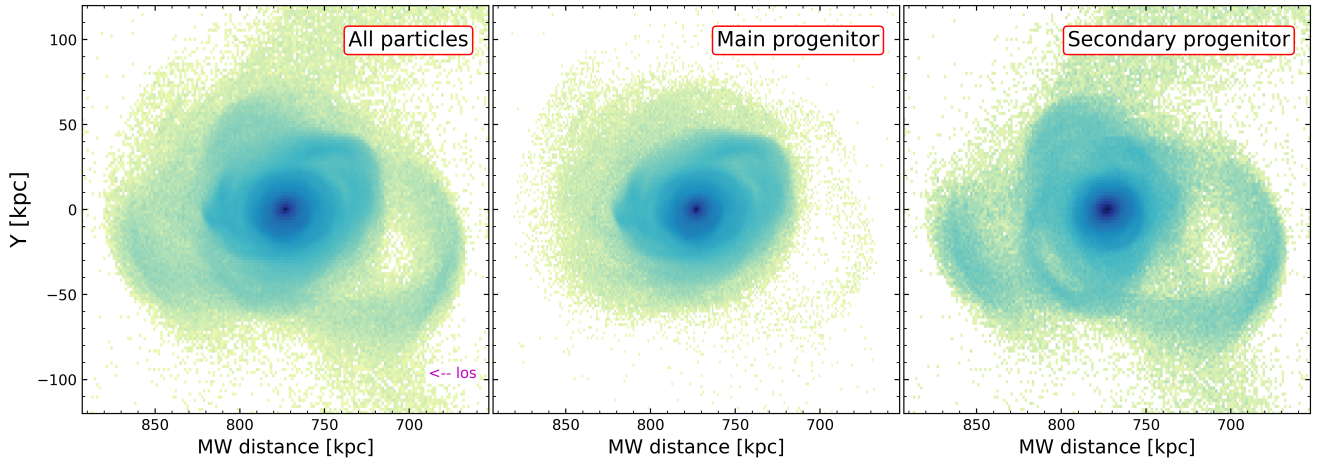
Fig. 5. Two-dimensional histogram of the LOS (Z, Y) plane view of the M31 remnant. The centre of the galaxy is fixed at 773 kpc from the MW (Conn et al. 2016). All simulated particles [left], main progenitor particles [middle], and secondary progenitor particles [right] are plotted separately. The Z-axis of the simulation denotes our LOS and is designated as MW distance (our vantage point is marked on the bottom right of the leftmost panel). The Y-axis corresponds to the distance from the centre of M31 in the direction from north (up) to south (down) in units of kiloparsecs.

In a major merger, the substructures acquire complex configurations in 3D space as a consequence of subsequent pericentric passages. The released tidal debris stars in the distinct loops produce over-densities and ridges in the observed phase space and complex structures along the LOS distance (H18). This is particularly interesting given the presence of the KCCs in the GSS (Kalirai et al. 2006; Gilbert et al. 2009), recently confirmed by the DESI observations of the inner halo of M31 (D23). In addition to the KCCs, the distributions of the LOS distances versus photometric metallicity estimates, from the resolved stellar populations studies carried out by Conn et al. (2016) and Ogami et al. (2025), show broad probability distributions for the LOS distances along the GSS with secondary-peaks, and σ[M/H]$\simeq$1 dex for the metallicity distributions in the GSS and inner halo substructures. See also the work by D23 for an illustration of the coherent kinematic structures (streams, wedges, and chevrons) in the positions and velocities distribution of individual stars at different locations in the M31 inner halo, as well as their metallicity distribution.

Using the H18 simulations, we can now explore the network along the LOS, phase space, and metallicity distributions of the GSS and NE and W shelves analogues in the M31 simulated remnant. In Figure 5, we show the LOS particle distribution of the simulated galaxy, plotted in the (Z, Y) plane, to illustrate its spatial properties. The composite nature of each structure emerges from the several loops seen in projection, mostly associated with the several pericentric passages along the infalling orbit of the secondary galaxy (see the rightmost panel in Figure 5). Their association with the GSS and NE and W shelves over-densities was previously presented in H18. Here we focus on their LOS distance structure, projected phase space properties, and metallicity distributions.

To group particles within each substructure of the inner halo of the simulated galaxy, we adopted Density-based Spatial Clustering of Applications with Noise (DBSCAN; Ester et al. 1996) which catalogues the multiple components within the GSS and NE and W shelves. This code utilises the 3D position of the stars rather than simply over-densities in the projected number density distribution on the sky. It recognises distinct clusters of points in high-density regions of any given parameter space. Our adopted

parameters and their specific implementation in the current study are described in detail in Appendix C.

Once particles are grouped in each of the three substructures (GSS and NE and W shelves) of the M31 simulated inner halo, we study the morphology of their $R_{\rm proj}$ versus $V_{\rm LOS}$ phase space diagrams, and their metallicity distribution. In Sect. 6, we shall carry out a one-to-one comparison of the [M/H] average values and standard deviations from relevant regions in the simulated M31 remnant with metallicity measurements for the same spatial regions of M31 itself.

In Figures 7, 9 and 11, we identify any stream-, chevron-, and wedge-like over-density patterns in each component of the resulting substructures in the simulated M31 remnant. We used the orange-dashed and black-dotted lines for ridge and shell pattern identification in the main and secondary particles, respectively. Whenever the same feature arose in the phase space distributions of stellar particles originating from either progenitor, we plot the lines in red. These morphological features are to be compared with the observed ones in the $R_{\rm proj}$ versus $V_{\rm LOS}$ diagrams recently acquired from the DESI observations in the inner halo of M31 (D23).

In the phase space diagrams shown in the following sections, we plot the $(V_{\rm LOS} - V_{\rm sys})$ versus $R_{\rm proj}$, where $V_{\rm sys}$ is the systemic velocity of M31 equal to $-300 \, \rm km \, s^{-1}$ (Watkins et al. 2013) and $R_{\rm proj}$ is the projected linear distance (in kpc) from the centre of the simulated M31 remnant, computed as $R_{\rm proj} = \sqrt{(X^2 + Y^2)}$.

4.1. The Giant Stellar Stream

The PAndAS survey (McConnachie et al. 2009) showed that the GSS is the dominant contributor to the number of RGB stars in the M31 inner halo. It covers at least $\sim$6° on the sky, and it reaches out to $\sim$80 kpc (McConnachie et al. 2003; Conn et al. 2016; Cohen et al. 2018) in projected distance. The proximity of M31 allows observers to study the resolved RGB population in the stream (e.g. Ibata et al. 2004; Guhathakurta et al. 2006; Ferguson & Mackey 2016; Wojno et al. 2023), and use them to estimate the LOS distances from the tip of the RGB (McConnachie et al. 2003; Conn et al. 2016; Ogami et al. 2025), in addition to obtaining the radial velocities of RGB stars (e.g.
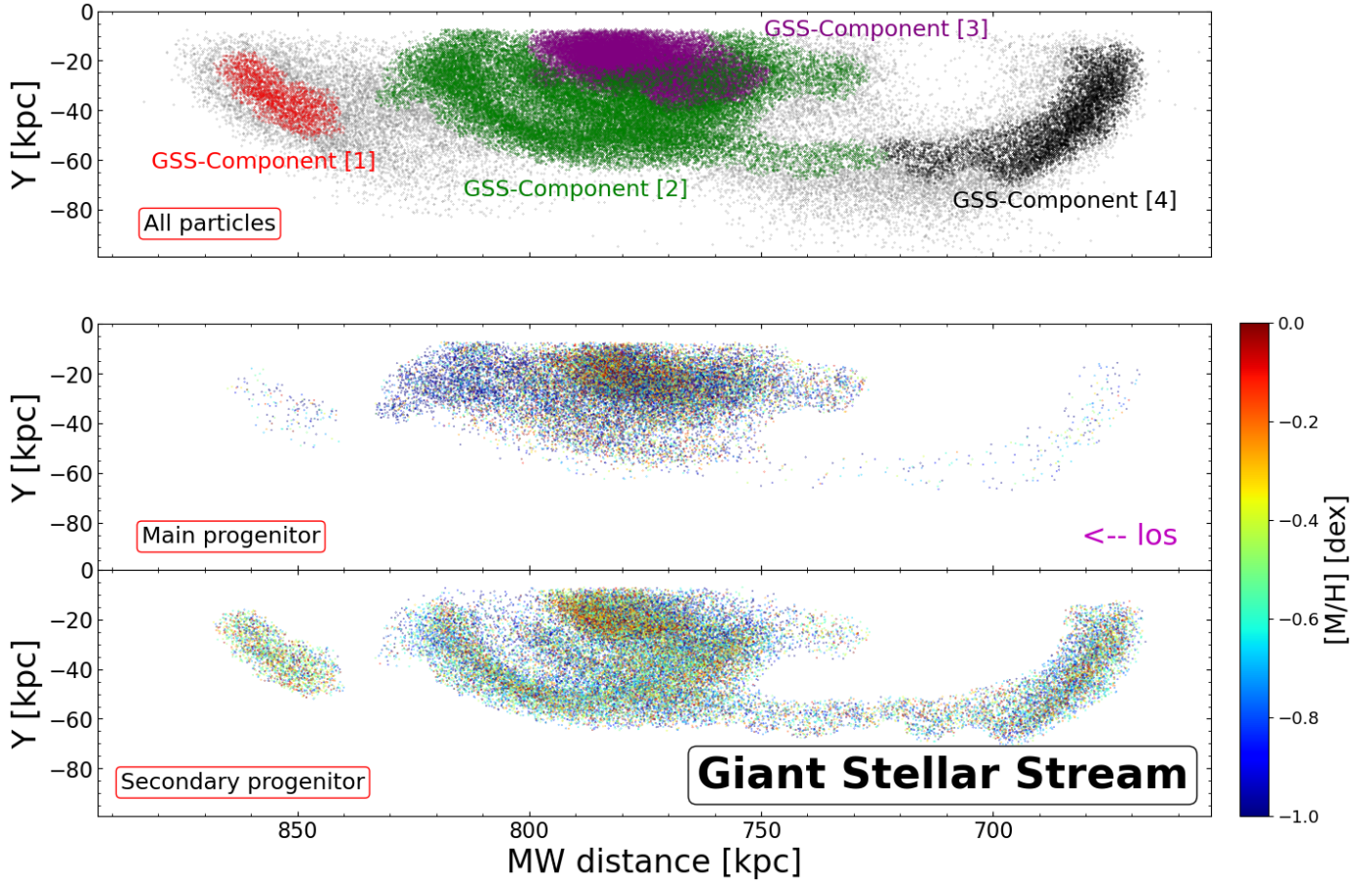
Fig. 6. Upper panel: Illustration of the distribution along the LOS of all the stellar particles of the spatially selected GSS (see Figure 1), grouped into four numbered components by DBSCAN. Grey points are stellar particles that are not associated with any group of particles according to the DBSCAN clustering algorithm. The entire spatial distribution of the modelled GSS stars along the LOS is compared to independent observations in Figure 14. Particles in the upper panel are colour-coded according to their association with a DBSCAN component, as detailed in the legend. Middle panel: Particles originating from the main progenitor that belong to the identified components in the GSS. They are colour-coded by their metallicity. The grey particles shown in the upper panel are not included. Lower panel: Same as the middle panel, but for stellar particles which originate from the secondary progenitor.

Guhathakurta et al. 2006; Gilbert et al. 2009; D23) and PNe (Bhattacharya et al. 2023).

In Figure 6, we show the (Z, Y) density distribution of the stellar particles in the region of the GSS, that is all the star particles extracted from the region enclosed by the red-dashed polygon of Figure 1 in the projected (X, Y) view on the sky of the M31 simulated remnant galaxy. We plot the star particles in the GSS region colour-coded according to the grouping identified by DBSCAN in the 3D particle distribution (X, Y, Z) of the M31 remnant. DBSCAN identifies four components of the GSS regions; they are labelled GSS-components [1] to [4] in the uppermost panel of Figure 6. In the middle and lower panels of Figure 6, we show the particles in groups [1] to [4] according to their origin, either main (middle panel) or secondary (lower panel).

The distribution in the (Z, Y) plane and the DBSCAN grouping clearly show that the S-components [1], [2], and [4] include the loops farthest away from the center. These loops are dominated by star particles originating from the secondary progenitor. The plane that contains the loops is seen nearly edge-on and the extended GSS structure is then the result of the superposition of GSS-components [1] to [4] along the LOS. GSS-component [3] is dominated instead by stellar particles from the main progenitor and lies closer to the M31's disc.

In the middle and lower panels of Figure 6, star particles are colour-coded according to their metal content. There is a definite metallicity gradient as a function of radial distance along the simulated GSS, with the more metal-rich star particles found closer to the centre of M31. The lower panel shows that the more metal-rich particles in the secondary are found in the DBSCAN components at smaller distances. In the middle panel, the star particles in the GSS-component [3], which originate from the main progenitor, show a radial metallicity gradient as well, with more metal-rich star particles found at smaller galaxy radii.

During a radial merger event, stars from the less massive companion become gravitationally unbound, and tidal debris are deposited in a leading and a trailing tail (Ferguson et al. 2002; Hendel & Johnston 2015). As in any galactic disruption event, the least bound stars in the disrupted satellite are deposited at the largest distances in the host potential (e.g. Amorisco 2017). As stars approach the apocentre of their orbits in the deeper potential well, they reverse their path, which leads to an enhancement in stellar density at the apocentre (Merrifield & Kuijken 1998). Their morphology will resemble a shell, and a V-shaped, chevron- or wedge-like feature will appear in their R_{proj} versus V_{LOS} diagram.
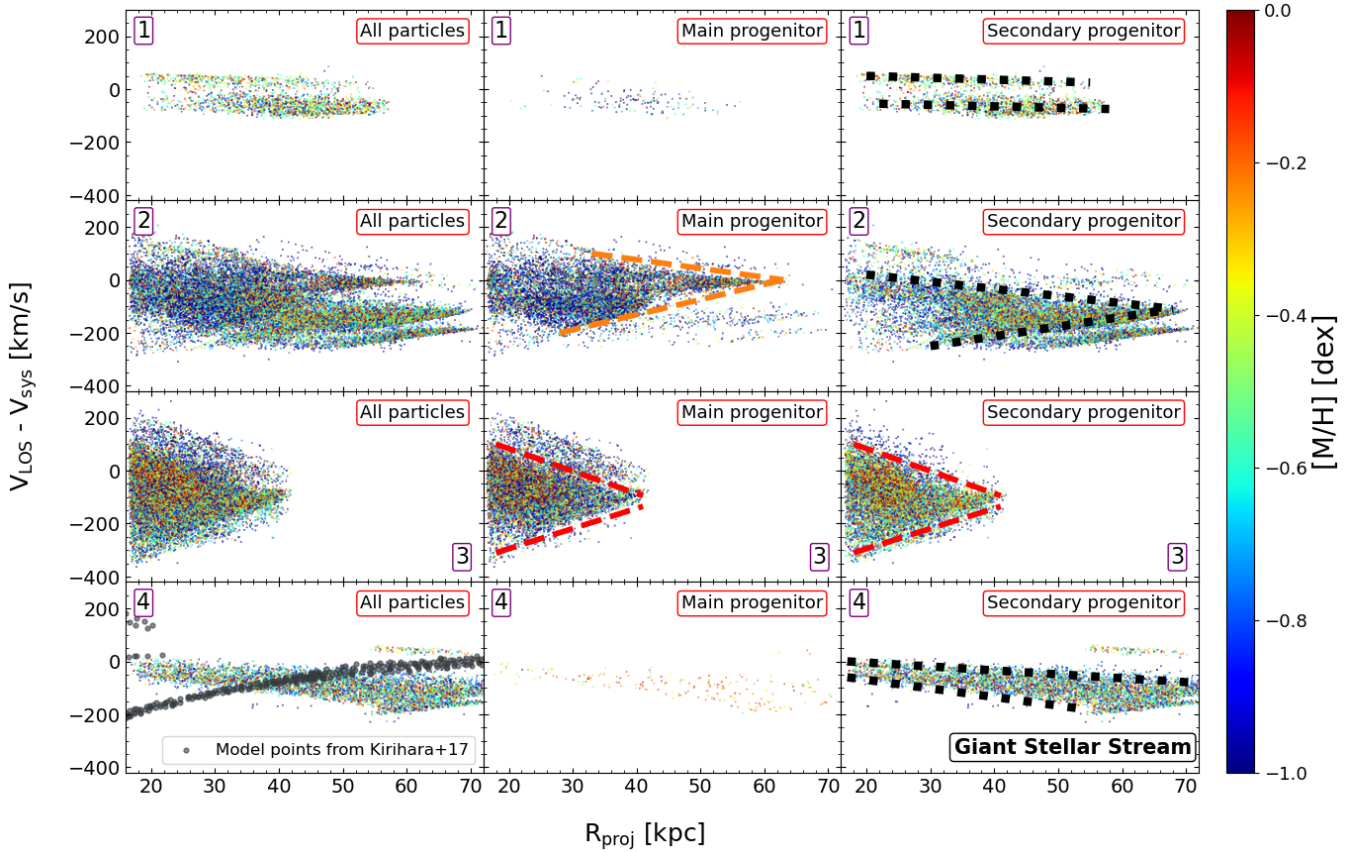
Fig. 7. Phase space diagram ($V_{LOS} - V_{sys}$) versus R_{proj}. Here, V_{sys} is the systemic velocity of M31 equal to $-300 \, \mathrm{km \, s^{-1}}$ (Watkins et al. 2013), and R_{proj} is the projected linear distance in kiloparsecs to the centre of the simulated M31 remnant for the modelled stars in the GSS. The four panels show the projected phase space for each GSS-component ([1] to [4]; see Figure 6) identified by DBSCAN. All the stellar particles are colour-coded by their metallicity. All simulated particles [left column], main progenitor particles [middle column], and secondary progenitor particles [right column] are separately illustrated. Ridges from the main progenitor are marked with orange-dashed lines while those from the secondary are marked with black-dotted lines. When the same morphological feature is identified in both progenitors, then it is marked by red-dashed lines in both panels. In the leftmost panel of the GSS-component [4] row, the grey circles are simulated stars from the satellite's trailing tail, reproduced from Kirihara et al. (2017).

The R_{proj} versus V_{LOS} phase space diagrams for all the star particles for the GSS-components [1] to [4] are plotted in Figure 7, leftmost panels. We illustrate the different regions of the star particle distribution in panels [1] to [4], and then we further showcase whether star particles originate from the main (central panel) or secondary (right panel) progenitor. We outline wedge patterns in GSS-components [2] and [3]. Streams and tail features appear in GSS-components [1] and [4], which are populated by stars originating from the secondary progenitor. The morphology of the GSS-components arises since the orbit of the centre of mass of the secondary progenitor becomes more radial as it sinks to the centre of the deeper potential well (as described in Amorisco 2017). Therefore, stellar debris removed during the early stages of the merger produces streams/arms (i.e. like GSS-components [1] and [4]), while stellar debris released later on generates wedges and chevrons (i.e. GSS-components [2] and [3]).

Specifically, the morphology of the phase space distribution of the GSS-component [1] displays a stream-like feature populated by particles from the secondary progenitor (Figure 7, panel [1]).

The phase space distribution for the GSS-component [2] displays two distinct V-shaped chevrons, populated by star particles from the main and the secondary particles (Figure 7, panel [2]).

These two chevrons reach different projected distances, with the chevron populated by the secondary star particles reaching larger distances and negative velocities, with its turnaround point at $V_{LOS} - V_{sys} \simeq 100 \, \mathrm{km \, s^{-1}}$ at $R_{proj} \simeq 70 \, \mathrm{kpc}$. Overall, the GSS-component [2] is a prominent feature of the simulated GSS. A wedge pattern is seen in the phase space distribution of the GSS-component [3] with similar density of star particles and ranges in radius and velocities for both progenitors; this feature is shown with a red dashed line to emphasise that we deal with the same feature in both progenitors (Figure 7, panel [3]). The GSS-component [4] is a tail/stream extended over large projected distances from the main body of the galaxy, towards the MW along the LOS. This tidal tail exhibits the largest projected velocity and distance from the centre of M31 among the GSS components. A stream-like feature is seen in its phase space distribution (Figure 7, panel [4]).

To summarise, the major merger simulations would lead to the presence of distinct kinematic components in the GSS region, which were previously identified in observations (Kalirai et al. 2006; Gilbert et al. 2009; Dey et al. 2023b). Such stars originate from either the main or the secondary progenitor (Bhattacharya et al. 2023). Hence, in the H18 merger simulation, the GSS appears to be a composite structure comprising loops of stars at different radial distances/energies, with
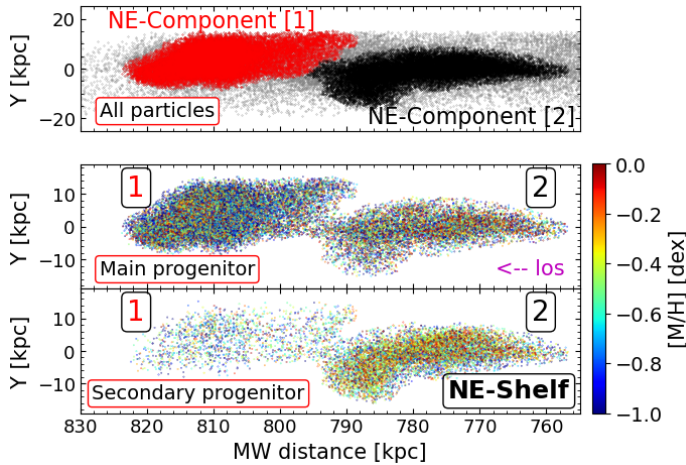
Fig. 8. Illustration of the distribution along the LOS of all the stellar particles of the spatially selected NE shelf (see Figure 1) grouped into two numbered components by DBSCAN (upper panel: red and black points labelled in the legend). Grey points are stellar particles that are not linked to any group of particles according to the DBSCAN clustering algorithm. The entire spatial distribution of the modelled NE shelf stars along the LOS is compared to independent observations in Figure 15. Middle and lower panels: Same as Figure 6 but for the particles in the NE shelf region.

significant structural differences from the smooth trailing stream of a minor accretion event. The complex kinematic structure of the GSS can only be reproduced by a major merger, since a minor merger would have produced a single trailing tail; see the grey points of Kirihara et al. (2017) simulation reproduced in the leftmost panel of Figure 7, GSS-component [4], all particles.

4.2. The North-East shelf

The NE shelf indicates a region exhibiting an abrupt increase in surface stellar density in the north-east region of the inner halo of M31 (Ferguson et al. 2002). The NE shelf is the outermost shell according to Escala et al. (2022), extending to $R_{proj} \simeq 31$ kpc and spanning ~ 400 km s^{-1} in velocity.

DBSCAN is used to identify clusters in the (X, Y, Z) particle distribution at the NE location in the M31 remnant that emerge as stellar overdensities in their 3D unfolding. Figure 8 reveals two components in the (Z, Y) plane of the selected region. These two clusters have different LOS distances to the MW, with NE-component [1] in the radial range of 820 to 800 kpc, and NE-component [2] in the 790 to 760 kpc range.

In Figures 8 and 9, we present the LOS structure of the NE shelf in the M31 remnant, and the R_{proj} versus V_{LOS} distribution for each component, respectively. In both figures, it is apparent that NE-component [1] comprises mostly particles from the main progenitor. In the R_{proj} versus V_{LOS} diagram of Figure 9, there is a single, extended feature in this component indicating that these stars exhibit disc kinematics. We also note that the star particles in component [1] show a wide spread of metallicity, with a large contribution from the range [M/H] < −0.6 dex. The assessment of the stellar ages within the rotationally supported NE-component [1] resulted in ∼23% of stars being younger than 5 Gyr old. Bernard et al. (2015) report that their "disc-like" fields in the NE shelf regions have had one-quarter of their mass formed in the last 5 Gyr. We then interpret component [1] as an extension of the rotationally supported M31 disc, whose younger

metal-poor stars are formed from the outermost gas in the main progenitor. For the contamination of the M31 disc stars in the NE shelf, see also the discussion in Escala et al. (2022).

Next, we turn to the NE-component [2]. This cluster of stars comprises wedges that are populated by stars originating either from the main or the secondary progenitor.

The leftmost panel of Figure 9 [lower row] featuring all the particles within NE-component [2], gives a vivid illustration of the broad metallicity spread predicted in this study for the NE shelf, with distinctly higher metallicity than for the rotationally supported NE-component [1]. The cluster of stellar particles in NE-component [2] of the NE shelf can be decomposed into two wedge patterns (see Figure 9), with an inner wedge appearing mainly in the R_{proj} versus V_{LOS} diagram of particles from the main progenitor, and an outer wedge populated by particles from the secondary. The lower part of the inner wedge appears in both distributions, so we show it as a red-dashed contour in both panels. The radial extent of the wedge in the NE shelf populated by particles from the secondary covers a range of radii and velocities very similar to the ranges for these quantities determined from the wedge distribution of the NE shelf RGB stars published by Escala et al. (2022).

We note that a difference in the maximum R_{proj} of the wedges suggests distinct apocentre points for the underlying coherently moving group of stars. Stars from the pre-merger disc (which have an apocentre closer to the centre of M31) slow to a halt before reversing their path sooner than particles from the secondary. This is expected since stars with different initial orbits reach distinct apocentre radii (Merrifield & Kuijken 1998).

A possible cause for NE-component [2] to include particles from the main is that, as the satellite passes through the pre-merger disc of the galaxy, stars are dragged along the orbits of the secondary. Therefore, finding particles in each stellar substructure in the inner halo coming from the pre-merger disc of M31 is expected in a major merger scenario, due to the substantial gravitational pull of the more massive secondary galaxy.

4.3. The Western shelf

The W shelf in M31 is a stellar overdensity signaled by an increased surface brightness, embedded in the low-surface brightness, smooth halo. It extends from the western side of M31's disc (Fardal et al. 2007; Ferguson & Mackey 2016).

Using DBSCAN, we identify two clusters in the (X, Y, Z) distribution of particles in the western region of the M31 remnant: W-component [1] and W-component [2], shown in Figure 10 [upper panel] in different colours. These two clusters have different LOS distances to the MW, with W-component [1] in the radial range of 770 to 750 kpc, and W-component [2] in the 750 to 730 kpc range. Both progenitors contribute to these clusters, with the secondary being more prominent in W-component [1], and the main progenitor in W-component [2]. In Figure 10, star particles in the middle and lower panels are colour-coded according to their metallicity; their coloured distribution shows that star particles from the secondary in W-component [1] have a higher metal content than those in W-component [2] originating from the main progenitor.

In Figure 11, we show the phase space distributions of the stellar particles in the W shelf components [1] and [2]. Two wedge-like patterns are evident in the W-component [1] (see Figure 11, upper panels); one is populated by particles from the main progenitor, and another pattern appears in the distribution of particles from the secondary. The two wedges span almost the same projected distance ($R_{proj,max} \simeq 26$ kpc) and velocity ranges.
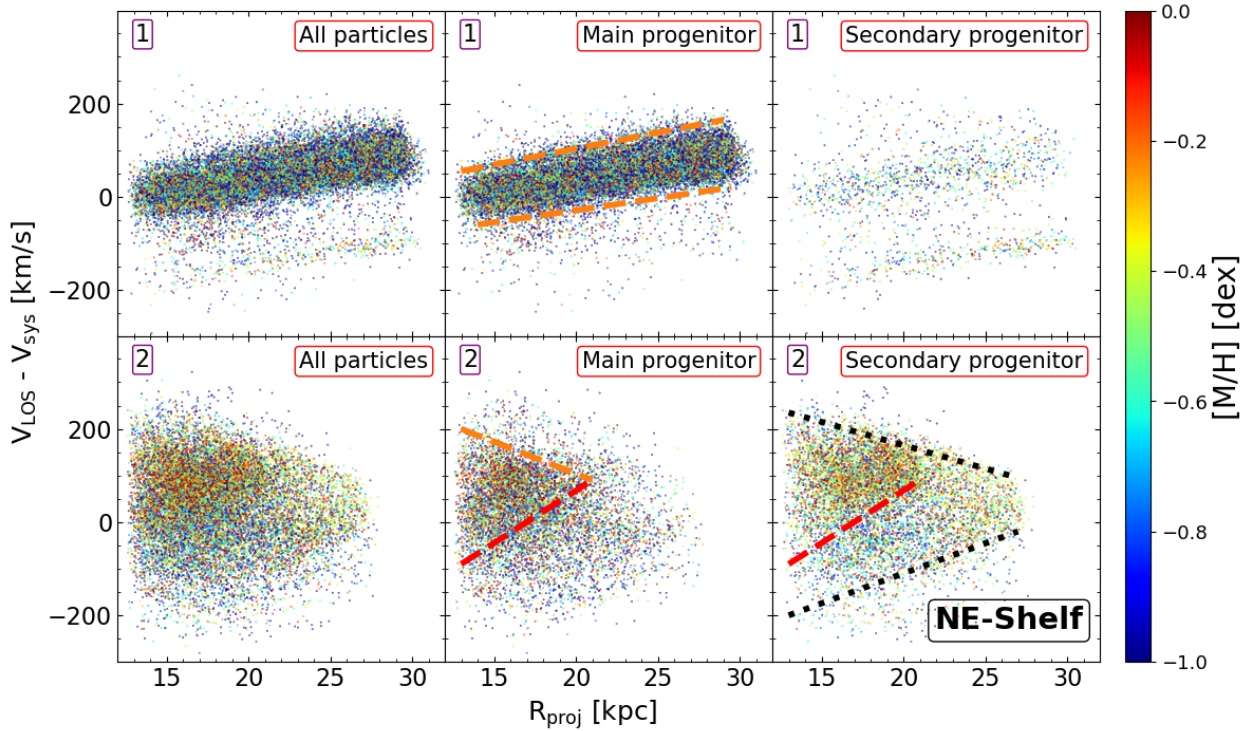
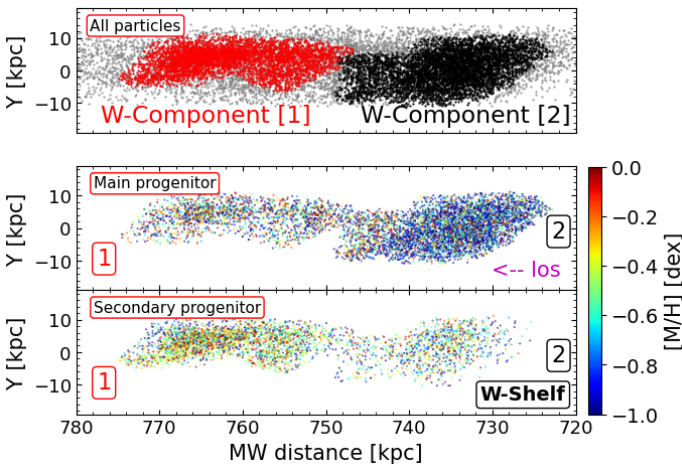**Fig. 9.** Same as Figure 7 but for the NE shelf with two components.

Fig. 10. Illustration of the distribution along the LOS of all the stellar particles of the spatially selected W shelf (see Figure 1), grouped into two numbered components by DBSCAN (upper panel, red and black points labelled in the legend). Grey points are stellar particles that are not linked to any group of particles according to the DBSCAN clustering algorithm. The entire spatial distribution of the modelled W shelf stars along the LOS is compared to independent observations in Figure 16. Middle and lower panel: Same as Figure 6 but for the W shelf.

A stream-like patch of stars is also identified in particles from the secondary. The presence of wedges from main and secondary stars and their colour-coded metallicity indicate a large width of the metal abundance distribution for this component as well.

The phase space distribution of the stellar particles in the W-component [2] shows that these star particles (coming from the pre-merger disc) exhibit disc-like kinematics, reaching a negative V_{LOS} consistent with the rotation of the M31 disc on its

western side and have lower metallicity than the stellar particles populating W-component [1].

5. R_{proj} versus V_{LOS} diagrams for the Giant Stellar Stream and the North-East and Western shelves from the 1:4 mass merger simulation

The simulations by H18 are optimised to reproduce numerous observed properties of M31's disc and the morphology of its inner halo. Our investigation aims at extending the comparison with M31 to the phase space and metallicity distributions of the halo substructures, including the recently acquired chemodynamical data from the wide-field spectroscopic sample of D23.

In this comparison, it is important to take into account the principal limitations of the model (see also Appendix D): (i) The H18 simulation is viewed at a specific snapshot during an ongoing evolution; (ii) while a large number of initial parameters was sampled to optimise previously available observations, the sampling is still coarse and important assumptions had to be made; (iii) the final viewing direction also has remaining uncertainties which influence the V_{LOS} towards the MW; (iv) the chemical model is based on a simplified enrichment scheme during the simulation, as well as on a posteriori comparison with limited data. Since the substructures in the halo of M31 are inherently time-dependent, these uncertainties can have significant effects: features seen in the data may not all be simultaneously found in the model. Consequently, we would not expect high levels of accuracy in comparing observed phase space features and their metallicity properties quantitatively with similar features in the model. It is therefore remarkable that multiple features seen in the data are reproduced qualitatively, and in some cases even quantitatively.

Over and under densities in the R_{proj} versus V_{LOS} diagram of the GSS stars were evident from previous spectroscopic surveys
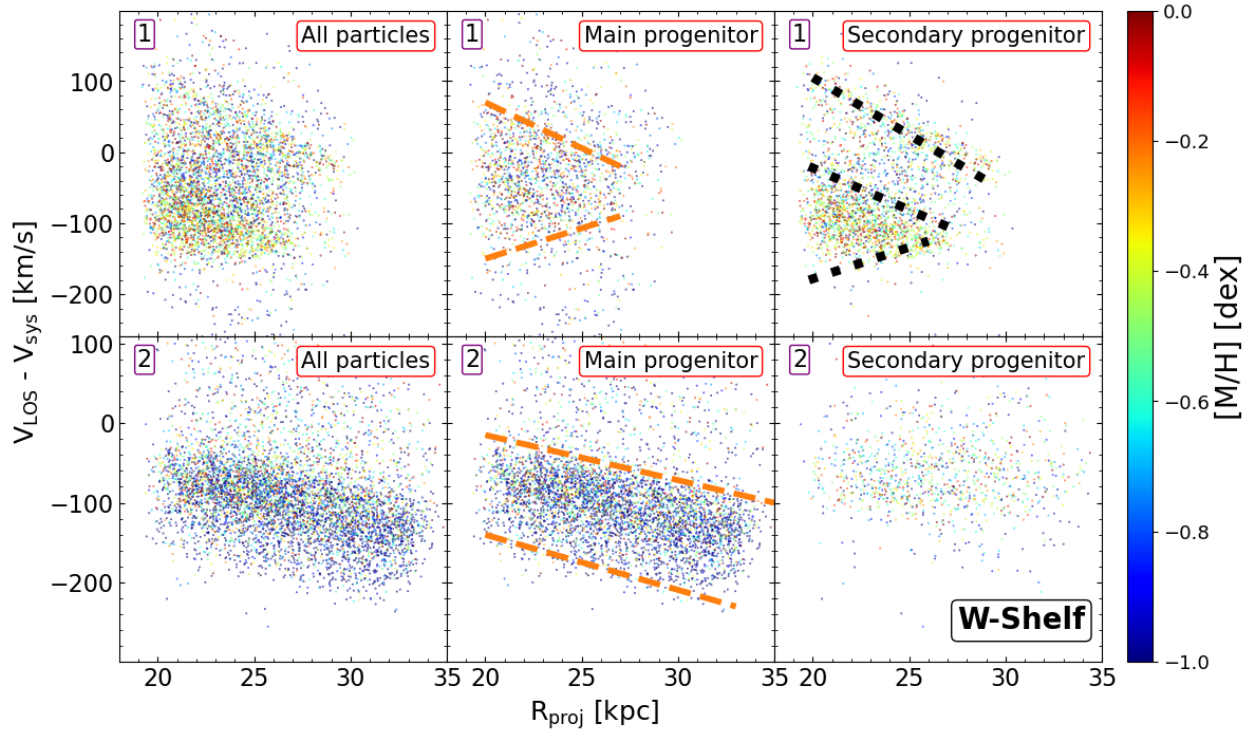
Fig. 11. Same as Figure 7, but for the W shelf with two components.

(Merrett et al. 2006; Gilbert et al. 2007, 2009; Fardal et al. 2012; Escala et al. 2020, 2022; Bhattacharya et al. 2023), and were already reproduced in simulations with different levels of success (e.g. Fardal et al. 2012; Kirihara et al. 2017; Milošević et al. 2022, 2024). Recently, the DESI survey (D23) provided a catalogue of ∼7000 sources residing in the inner halo of M31, selected according to criteria shown in their Figure 1, with measured radial velocities and projected distances, along with spectroscopic metallicity estimates. This is the most uniform spectroscopic coverage of the inner halo of M31 to date.

In Figure 12, we reproduce the R_{proj} versus $(V_{LOS} - V_{sys})$ distributions from the DESI survey (grey points) extracted in the regions co-spatial with GSS and NE and W shelves. On these distributions, we overplot the R_{proj} and $(V_{LOS} - V_{sys})$ measurements for the PNe in the same regions as red and green stars, from Merrett et al. (2003, 2006), and Bhattacharya et al. (2023), respectively. To guide the eye, the wedge features already identified and labelled in D23 are reproduced in the plots. They are colour- and line-coded according to similar features in the M31 model specified in Sect. 4, Figures 7, 9, 11. Some of the overdensities identified here are not labelled or identified in D23. Specifically, the lower envelopes of the two wedge-like patterns that emerge in the phase space diagram of the NE shelf are not labelled in D23 but are identified as such in this study. We discuss the comparison between the observed and the simulated phase space for each substructure in the inner halo of M31 in turn in the following sections.

5.1. Giant Stellar Stream phase space comparison

In the leftmost panel of Figure 12, we examine the DESI RGBs (D23) and PNe (Bhattacharya et al. 2023) phase space at the location of the GSS. D23 identified a chevron at a low projected distance of $R_{proj} \leq 20\,kpc$ (the red-dashed lines in Figure 12 labelled "1cr, 1cb"). A similar chevron structure is evident in the

simulated GSS (see the phase space of GSS-component [3] in Figure 7), with the same morphology for stellar particles from both progenitors. The tip of the wedge occurs at $R_{proj} \simeq 40\,kpc$ in the simulation model.

Another feature identified in the observed GSS phase space as '1ab' by D23, was initially reported by Gilbert et al. (2009) to be a kinematically cold component, that is a stream of stars whose velocity varies between ∼−300 km s⁻¹ for R_{proj} ∼ 10 kpc to ∼−100 km s⁻¹ for R_{proj} ∼ 55 kpc. In Figure 7, we find an elongated wedge pattern in the secondary particles of GSS-component [2], whose lower edge closely resembles the observed feature labelled '1ab' in the DESI phase space.

In addition, there is the overdensity labelled '1bb' by D23. We note that '1bb' and '1ab' are the counterparts in the DESI phase space of the GSS KCC (Kalirai et al. 2006; Gilbert et al. 2007). When examining the simulated phase space in Figure 7, the lower edges of the distributions of the main and secondary stellar particles in the GSS-component [2] are reminiscent of these features, as already noted in Bhattacharya et al. (2023).

We are guided by a similarity in R_{proj} and V_{LOS} ranges to associate GSS-component [2] and [3] with the prominent overdensity features of the DESI phase space diagram for the RGBs in the GSS. Differently, we do not find a matching structure for GSS-component [1] and [4], seen in Figure 7. We acknowledge that GSS-component [1] exhibits low surface brightness; therefore, it may not be detectable. The GSS-component [4] is placed at a particularly large LOS distance and is most affected by projection effects. Furthermore, as discussed in H18, a small shift in the initial conditions strongly affects its kinematics[5].

---

[5] The specific temporal choice made in model #336 for the entire $z = 0$ snapshot is also relevant. If one allows the model to evolve further in time, it leads to the accretion of an additional loop surrounding the post-merger disc, while in reality, it has not reached the galaxy yet.
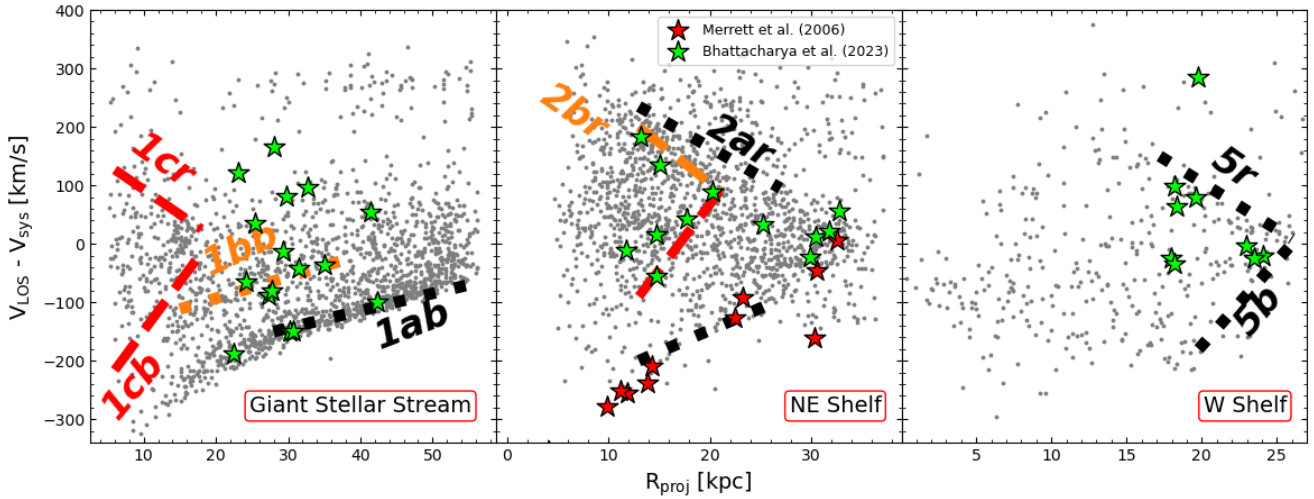
Fig. 12. Position versus ($V_{\text{LOS}} - V_{\text{sys}}$) velocity diagrams for three areas in the DESI survey. Each grey point is a star identified in the DESI survey. Each selected area encloses the substructure denoted by the bottom right label (GSS [left], NE shelf [middle], and W shelf [right]). Linear features (streams, wedges, and chevrons) from DESI are overplotted, with the labels used by D23. Orange dashed and black dotted lines designate particles from the main and the secondary progenitor, respectively, of analogous model features in Figs. 7, 9, 11. Whenever the same feature appears in the phase space of stars from both progenitors, it is sketched with red-dashed lines. The PNe data from Bhattacharya et al. (2023) are overplotted for each substructure as green stars. The PNe, classified as outliers by Merrett et al. (2003, 2006), are overplotted in the NE shelf region as red stars.

In summary, we find that wedges, chevrons, and KCCs in the GSS can be produced by several wraps of the merging secondary galaxy in the model. A moderate contribution from the main progenitor stars along the orbit of the merging secondary is also expected.

5.2. North-East shelf phase space comparison

We next examine the R_{proj} versus V_{LOS} diagram of stars in the DESI survey at the location of the NE shelf; see the middle panel of Figure 12. D23 identified as over-densities the linear structures labelled '2br' and '2ar'. We also note that the edge at negative velocities in the radial range of R_{proj} from 10 to 30 kpc was prominent in the NE shelf phase space in Figure 10 of Escala et al. (2022), and is populated by the PN outliers in Merrett et al. (2003, 2006). In Figure 12 [middle], PNe data from Bhattacharya et al. (2023) are plotted as green stars, while PNe outliers from Merrett et al. (2003, 2006) are plotted as red stars.

In Figure 9, the phase space of the simulated NE-component [2] showed clearly two V-shaped patterns: an inner wedge populated by particles from the main galaxy, and an outer wedge populated by particles from the secondary. The inner chevron of NE-component [2] is strongly reminiscent of the feature '2br' in D23 (orange-dotted line in Fig. 12). The inner wedge in the simulated NE shelf matches the observed one in both the R_{proj} and V_{LOS} positions.

To reproduce the outermost wedge pattern that appears in DESI phase space for the NE shelf, with its tip at $\sim$34 kpc, we demarcate on Figure 12 the upper and lower black dotted lines from secondary particles of the NE-component [2]. The upper line traces the upper envelope of the observed wedge, while the lower line traces the corresponding observed lower envelope.

We also note here that the NE shelf phase space predicted from N-body models simulating a minor merger event predicts overdensities in the NE shelf suggestive of an intact core of the secondary progenitor (Fardal et al. 2007; Escala et al. 2022). Such overdensities are morphologically very different from the inner wedge associated with '2br' in D23. The double wedge

structure emerges in H18 simulations at the location of the NE shelf as particles from the main progenitor are pulled by the passage of the secondary. A sufficiently massive object is required. Hence, the agreement with the DESI features of the chevron- and wedge-like overdensities in the projected phase space diagram for the NE shelf gives additional support to the major merger scenario for the origin of the M31 inner halo substructures.

5.3. Western shelf phase space comparison

In the DESI phase space at the location of the W shelf (Figure 12 [right]), the labelled overdensities are '5r' and '5b', delimiting the upper and lower envelopes of the DESI phase space wedge distribution. The linear outer edges in the W-component [1] in the upper panels of Figure 11 are reminiscent of the wedge in the DESI phase space, which extends over the very same radial and velocity ranges. The PN data in the W shelf region from Bhattacharya et al. (2023) (green stars) also trace the outer contour of the wedge pattern.

We note that while W-component [2] is clearly evident in Figure 11, it is not visible in the D23 observed phase space (Figure 12). This can be understood in terms of target selection bias in the DESI survey which required a metallicity of $[\text{M/H}] > -0.5$ dex (their Figure 1) while the W-component [2] is more metal poor ($[\text{M/H}] < -0.6$ dex).

In our model, the wedge at the location of the W shelf is populated by star particles from both main and secondary progenitors. This property has implications for the abundance distribution in the W shelf. This holds also for the GSS and NE shelf, where similar superpositions along the LOS occur. This topic is investigated further in Sect. 6.

As a final remark, we note that PNe seem to trace the densest regions delineating the chevrons and wedge patterns in the NE and W shelves, and in the GSS in Figure 12. PNe, being discrete tracers of light and stellar kinematics, allow the overdensities in phase space to be traced, which are the chevrons and wedge patterns in the M31 substructures in this work. A similar property

Table 3. Median metallicity ([M/H]) values and their standard deviation estimates from Escala et al. (2020), Escala et al. (2022), and D23, and the predicted values for the equivalent regions in the simulated M31 remnant.

Region	D23	Escala et al. (2020, 2022)	Component 1 [C1]		Component 2 [C2]		Component 3 [C3]		Component 4 [C4]	
			Main	Sec	Main	Sec	Main	Sec	Main	Sec
Giant Stellar Stream	-0.37 ± 0.72	-0.52 ± 0.45	-0.79 ± 0.52	-0.49 ± 0.31	-0.78 ± 0.59	-0.61 ± 0.36	-0.56 ± 0.50	-0.53 ± 0.38	-0.74 ± 0.37	-0.60 ± 0.33
NE Shelf	-0.35 ± 0.75	-0.44 ± 0.37	-0.65 ± 0.49	-0.60 ± 0.41	-0.50 ± 0.48	-0.45 ± 0.36	–	–	–	–
W Shelf	-0.43 ± 0.90	–	-0.56 ± 0.51	-0.45 ± 0.36	-0.89 ± 0.55	-0.55 ± 0.39	–	–	–	–

Notes. For Escala et al. (2020), we used Equation (1) from Bhattacharya et al. (2021) to convert [Fe/H] values into [M/H] taking into account the relative α-enhancement within the field. For Escala et al. (2022) we calculated the average photometric metallicity (and its standard deviation) of the six fields within the NE shelf. For modelled stars, the median and the standard deviation values for each component of the substructures are separately presented for main and secondary stellar particles.

is shown by the PN overdensity tracing the crown in M87, as discussed in Longobardi et al. (2015).

6. Abundance distributions in the inner halo substructures: Predictions and comparison with observations

We now turn to examining the median metallicity of each component of the three substructures in the model, distinguishing between main and secondary particles. We also compare with the median spectroscopic metallicity of the equivalent region in D23. For the GSS and the NE shelf, we include in our comparison the metallicity estimates from the SPLASH survey (Escala et al. 2020 and Escala et al. 2022[6], respectively). Table 3 and Figure 13 report available spectroscopic metallicity measurements and the corresponding quantities predicted for each component of the simulated M31 inner halo substructures. The error bars denote the standard deviation of the metallicity in the simulation and the observations.

We note that the selection method for M31 stars implemented in the DESI survey is biased towards redder sources in $(g - i)$ colour and therefore may primarily sample the metal-rich and older RGB, as well as the redder AGB stars; hence, they do not sample the metal-poor regions well (D23). Consequently, we expect our model-predicted metallicity values to be slightly metal-poorer on average.

Apart from pencil-beam spectroscopic measurements at specific locations, estimates of the metallicity from photometry for the resolved stellar populations in the GSS were carried out over an extended area. Using a two-fold (stellar magnitude and colour) algorithm implemented on the PAndAS resolved RGB photometric dataset (McConnachie et al. 2009), Conn et al. (2016) constrained distances from the MW through the TRGB method and photometric metallicities along the GSS. The measurements by Conn et al. (2016) show variations in LOS distances along the GSS, a variation in the [M/H] average values along the stream, with the most metal-rich section at 1 degree ($\simeq$40–50 kpc) along the stream, and a significant spread of σ[Fe/H], up to 1 dex, for the metallicity distribution of the RGB population of GSS stars.

More recently, Ogami et al. (2025) utilised a novel approach to account for the severe contamination from the MW's stellar halo[7] and obtained a sample of ~90% accuracy in select-

ing M31 stars. They provide the photometric metallicity distribution, projected distances from the M31 center, and distances from the MW. This survey covers the regions of the GSS, the NE and W shelves. Ogami et al. (2025) confirm the variation of [M/H] along the GSS, and a large σ[M/H] $\simeq$ 1 dex, which implies the presence of solar and super-solar metallicity stars in the GSS. Ogami et al. (2025) find that stars in the NE and W shelves have average values of their metallicity distribution of $\simeq$−0.5 dex with σ[M/H] $\simeq$0.7–0.9 dex, which are also significantly broad distributions. Photometric metallicity estimates for the resolved stellar population in the NE shelf are also reported in Escala et al. (2022) and Bernard et al. (2015), and for the W shelf in Tanaka et al. (2010). The metallicity distributions for each substructure are shown in Figures 14, 15, and 16, and a comparison with observations is reported below.

6.1. Line-of-sight distance and metallicity distributions for the Giant Stellar Stream

In Figure 13 [left] and Table 3, we report the metallicity values estimated for the GSS components, from the simulated major merger combined with a well-motivated chemical model. These values are generally consistent with the inferred spectroscopic metallicity data from Escala et al. (2020), while they are slightly metal-poorer than those from DESI, as anticipated based on their selection bias (D23). The widths of the metallicity distributions determined via spectroscopy of RGB stars in the GSS component by Escala et al. (2020) and D23 are narrower than those determined photometrically for the RGB population located in the GSS regions.

In the model, the lower metallicities of GSS-components [1] (main [M/H] $\simeq$ −0.8 dex, secondary [M/H] $\simeq$ −0.5) and [4] (main [M/H] $\simeq$ −0.74 dex, secondary [M/H] $\simeq$ −0.6, see Table 3) are consistent with being populated by stars located in the outermost parts of their progenitors' discs and, thus, with lower metallicity.

In Figure 14, we show the comparison between our simulation-inferred i) M31 LOS distance versus $R_{\rm proj}$, ii) mean metallicity [Fe/H] versus $R_{\rm proj}$, and iii) standard deviation of the metallicity σ[Fe/H] versus $R_{\rm proj}$ of stars in the GSS, and compare them with the equivalent quantities inferred from photometric measurements of resolved stellar populations in Conn et al. (2016) and Ogami et al. (2025)[8]. The metallicity values and their

---

[6] Escala et al. (2022) have acquired spectra for stars in the NE shelf region. However, they estimate the photometric metallicity after they make a kinematic selection to pin down NE shelf stars in order to minimise the contamination from the M31 disc.

[7] MW dwarf stars occupy the same region in the colour-magnitude diagram as the M31 giants.

[8] The reader should bear in mind that the stellar ages assumed in the independent photometric estimates of the metallicity distribution of the stars in GSS and NE and W shelves are different; Conn et al. (2016) assume a fixed age of 9 Gyr for the GSS, while Ogami et al. (2025) and Escala et al. (2022) assume a fixed age of 13 Gyr and 12 Gyr respectively. This critical assumption breaks down in the GSS and the other substructures closest to the M31 disc; see for example the age spread of targets in Bernard et al. (2015).
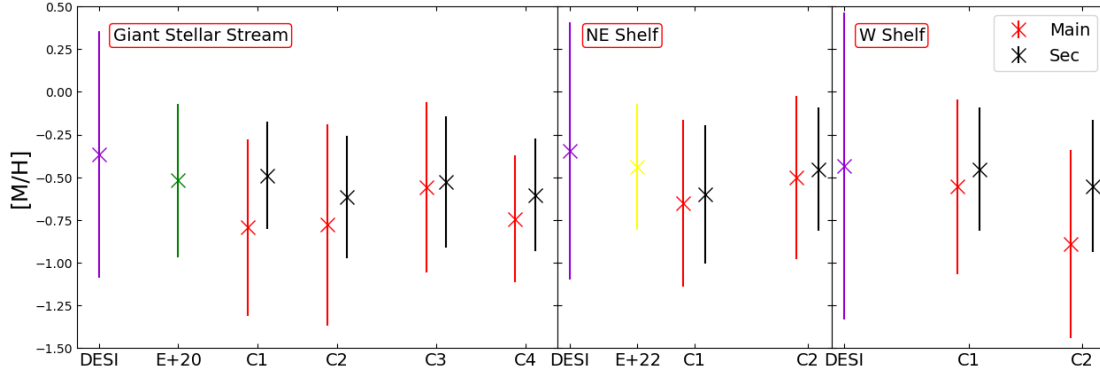
Fig. 13. Summary overview of the median metallicity values within GSS, NE and W shelves of the model to be compared with spectroscopic metallicity estimates from the DESI survey, Escala et al. 2020 (for the GSS; labelled as E+20 in the figure), and Escala et al. (2022) (for the NE shelf; labelled as E+22 in the figure). For each substructure, we report the median values from their components identified in the simulations by DBSCAN and labelled as C1, C2, etc. The error bars denote the standard deviation both for the data distributions and the distributions obtained from the simulation model.
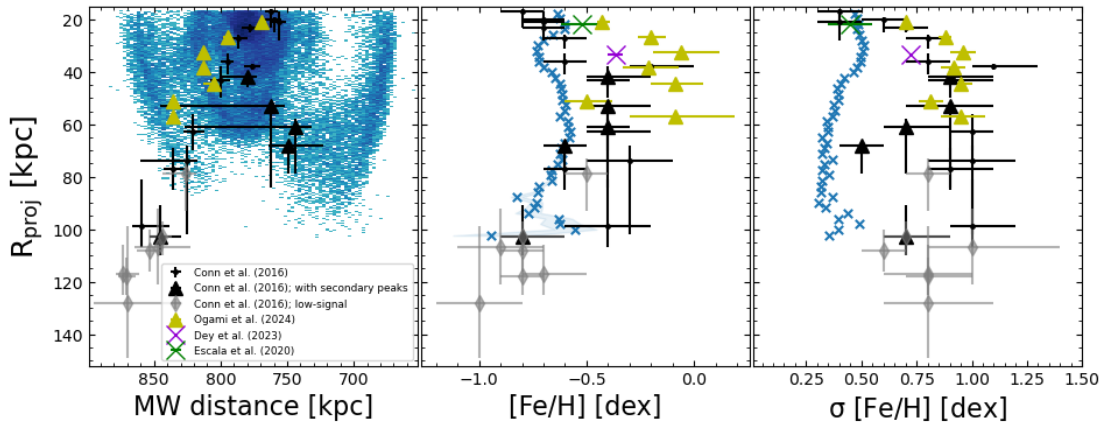
Fig. 14. Left panel: M31 R_{proj} and LOS distance for various regions in the GSS, from photometric measurements of Conn et al. (2016) and Ogami et al. (2025). For Conn et al. (2016) GSS sub-fields, small black circles denote distances and metallicities inferred from the single highest peak of their PDF, black triangles for quantities inferred by taking into account a secondary peak in the PDF, and grey diamonds for the fields with low signal-to-noise and high contamination fraction. Ogami et al. (2025) data are shown as yellow triangles. In blue we show the surface number density of star particles in the model. Middle panel: Comparison of the GSS's average metallicity values from Conn et al. (2016), Escala et al. (2020), D23 and Ogami et al. (2025) with the equivalent values predicted by the model (indicated by blue crosses), binned in ~2 kpc bins for clarity. The error bars for the [M/H] estimated by Ogami et al. (2025) represent the 68% uncertainty. Ogami et al. (2025) estimate their total systematic error in the estimated MW distance to be ~27 kpc. Error bars for the data of Conn et al. (2016) represent their 1σ (68.2 percent) uncertainties. Conn et al. (2016) assume a fixed age of 9 Gyr for the GSS, while Ogami et al. (2025) assume a fixed age of 13 Gyr. The critical assumption of constant age breaks down in the GSS part which is close to the M31 disc (see for example the age spread in the resolved stellar population in Bernard et al. 2015). Right panel: Comparison of the σ[Fe/H] for stars in the model with equivalent measurements from observed datasets of Conn et al. (2016), Escala et al. (2020), D23 and Ogami et al. (2025).

standard deviation from Escala et al. (2020) and D23 are also included. The modelled stars selected and plotted in Figure 14 included all stars in the spatial selection of the GSS in Figure 1 along the entire LOS.

Conn et al. (2016) divide the GSS into spatially adjacent sub-fields and estimate the probability distribution function (PDF) for the heliocentric distance, the average photometric metallicity [M/H], and the RGB metallicity spread of each field. For some of the fields that comprise the GSS, they found double peaks in the PDFs. These are independent evidence for several distinct components in the GSS along the LOS. In Figure 14, the larger, black triangles show the quantities when secondary peaks of the PDFs are present. We further comment that the agreement of the model's estimation of the LOS distance for stars in the GSS is sufficiently good out to an $R_{proj} \sim 80$ kpc. The outermost fields observed by Conn et al. (2016) are plotted in grey, due to the low

signal available and a very high contamination fraction (~90%) from foreground dwarf stars in the MW halo in those regions. Contrary to the Conn et al. (2016) data, the model GSS does not extend to $R_{proj} > 80$ kpc. As discussed already, some discrepancies are expected, as the major merger simulation was not tailored to reproduce these observations. We note that H18 feature an alternative model (#255) that reproduces the 3D structure of the GSS more closely. This model features a shorter radial distance for the first pericentre passage of the two progenitor galaxies, which defines the time elapsed between two passages. Since GSS comprises tidal debris that is retracted to the gravitational potential of the remnant after the merger, the time between the passages determines its morphology.

An important success of the model is the reproduction of the metallicity trend along the GSS from Conn et al. (2016) observations, which features a kink at a projected distance

of ∼50–70 kpc. The model also predicts a broad RGB metal-licity distribution (≃0.3–0.5 dex). We discuss these in turn below.

The kink at $R_{proj} \simeq$ 40–50 kpc arises naturally in the major merger simulation as superposition along the LOS of star particles in wedges of different radial extensions from main and secondary stars. Hence, we do not need to assume a steep initial metallicity gradient (see Milošević et al. 2022, 2024, where an initial metallicity gradient of −0.3 dex/kpc is adopted for the secondary galaxy). The multiple superposed loops and wedges that build the GSS along the LOS in the simulation here lead to the observed metallicity patterns and curvature as a function of R_{proj}.

We also compare the average values of [M/H] and the metallicity spread (σ[M/H]) for GSS regions in the model with equivalent quantities from Ogami et al. (2025). Their photometric average metallicities are higher than those of the model and also of the measured values by Conn et al. (2016). A possible bias in the observed stars is discussed in Appendix E. Appendix E shows that a better agreement with the mean photometric metallicity in Ogami et al. (2025) can be reached by selecting the younger (<3.5 Gyr) stellar population of modelled stars.

In Figure 14 rightmost panel, we show the σ[M/H] from the RGB width estimated by Conn et al. (2016) (black), Ogami et al. (2025) (yellow), Escala et al. (2020) (green), and D23 (purple symbols) compared with the model predictions. While the model predicts a σ[M/H] in the range 0.3–0.6, the photometrically inferred σ[M/H] are larger, in the range 0.5–1.0 dex.

We note that the super-solar metallicities of the stars in the KCC of the GSS from D23, the high average metallicity values obtained by Ogami et al. (2025), and the significant width of the RGB population (∼1 dex) for a large region of the GSS (Conn et al. 2016; Ogami et al. 2025) all point to a massive progenitor galaxy with an extended star formation history. This is favourable for reaching such high values of stellar metallicity as per the MZR (Maiolino & Mannucci 2019) and producing a wide spread in LOS distances and metallicities.

6.2. Line-of-sight distances and metallicity distribution for the North-East shelf

In Figure 13 [middle] and Table 3, we list the estimated metallicity average values for the NE shelf in our simulation model. We note that for this comparison, the computed [M/H] and σ[M/H] are evaluated for NE-component [2] only, since also the spectroscopic and photometric metallicity estimates (Richardson et al. 2008; Bernard et al. 2015; Bhattacharya et al. 2021; Escala et al. 2022) are derived for the shelf population only, while the contributions of the M31 disc plus halo, and MW halo are removed in the observations. As expected, the average metallicity values from the model are slightly metal-poorer than D23's spectroscopic metallicity values; see Figure 13 [middle].

In Figure 15, we show the surface number density of stellar particles versus LOS distances [left], the average metallicity values in our model binned in 2 kpc bins [middle], and the σ[Fe/H] [right]. In these plots, we also show the photometric metallicity, LOS distances, and metallicity width estimates from Ogami et al. (2025), Bernard et al. (2015), and Escala et al. (2022).

In our model, the broad RGB width arises naturally from the superposition along the LOS of wedges in the NE shelf phase space, which are populated by main and secondary star particles (see NE shelf component[2] in Figure 9). The σ[Fe/H] of the model is in excellent agreement with the equivalent values
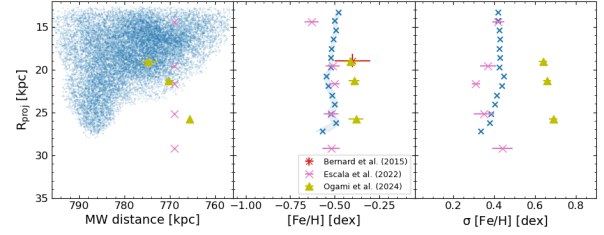

Fig. 15. Line-of-sight distribution, average metallicity and metallicity spread for NE shelf component[2], blue symbols. Ogami et al. (2025) data are shown as yellow triangles. The metallicity values and error bars from Escala et al. (2022) are shown in pink. Whenever they report two components in an observed field, their mean is plotted. The same MW distance of 769 kpc is assumed for all their NE shelf fields. Ogami et al. (2025) and Escala et al. (2022) assume a fixed age of 13 Gyr and 12 Gyr respectively. This critical assumption breaks down for distances near the M31 disc; see for example the age spread in Bernard et al. (2015). The Bernard et al. (2015) measurement is shown in red. For the projected distance, we have taken the average of the two fields for the NE shelf, while the relevant error corresponds to the size of each field.
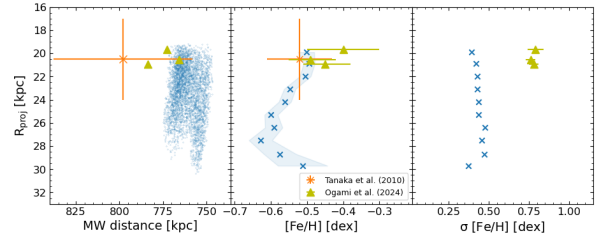

Fig. 16. Line-of-sight distribution, average metallicity and metallicity spread for W shelf component[1], blue symbols. Ogami et al. (2025) data are shown as yellow triangles. The plotted [M/H] and its relevant error from Tanaka et al. (2010) are obtained from their measured α and [Fe/H] abundances as described in Bhattacharya et al. (2021); they are presented in orange in the figures. For their R_{proj} we estimated the average of the projected distance of their surveyed fields, and the error was calculated as the distance of this average from the field edges.

for the NE shelf as measured by Escala et al. (2022). Instead Ogami et al. (2025) obtain a higher [M/H] average and larger σ[Fe/H] values, possibly from the contribution of other M31 components (e.g. the disc or the halo) superposed on the NE shelf.

6.3. Line-of-sight distances and metallicity distribution for the Western shelf

In Figure 13 [right] and Table 3, we show that our simulated metallicity values for the W-component [1] are consistent with the spectroscopic metallicity reported by the DESI survey for this substructure.

In Figure 16, we also report the predicted average [M/H] and RGB width (σ[M/H]) values for the simulated stellar particles for the W shelf component[1] in Figure 11, together with the metallicity average values, σ[M/H], and distance estimates from Tanaka et al. (2010) and Ogami et al. (2025). LOS distances and average [M/H] values agree very well with the observed values. The σ[M/H] values reported in Ogami et al. (2025) are ≃0.8 dex while in the simulation model, we obtain 0.3–0.5 dex. The larger spread in Ogami et al. (2025) may be related to superposed contributions of other M31 components like the disc and the halo, in addition to W shelf stars.

7. Conclusions

The giant Andromeda galaxy offers a unique opportunity for a close-up study of the hierarchical mass assembly of galaxies predicted by the ΛCDM model (White & Rees 1978; Bullock & Johnston 2005).

We utilised an N-body hydrodynamical simulation of a major merger between two disc galaxies with a mass ratio of 1:4 and combined it with a well-motivated chemical model to determine the chemodynamical properties of the M31 merger remnant. The combined model shows (i) a high metallicity and a large spread in the GSS and NE and W shelves, thus explaining the various photometric and spectroscopic measurements; (ii) a large distance spread in the GSS suggested by previous TRGB measurements; and (iii) phase space ridges, wedges and chevrons for all three substructures generated by several distinct pericentre passages of the secondary progenitor together with up-scattered main M31 disc stars, that also have plausible counterparts in the observed phase spaces.

These comparisons provide further independent and strong arguments for a major satellite merger in M31 ~3 Gyr ago and a coherent explanation for many observational results that make M31 appear so different from the MW. The wide spread in metallicities and LOS distances of the stars comprising the GSS serves as evidence for a complex and extended star formation history in the satellite galaxy and is suggestive of a relatively massive galaxy.

We compared the emerging chemodynamical properties of the substructures in the inner halo of the M31 simulated analogue, the GSS and the two shelves, with the available data in the literature. Our main results are as follows:

- We observed a multi-component nature for the NE and W shelves and especially for the GSS. In a major merger, the GSS contains multiple overlapping components along the LOS stemming from consecutive wraps of the satellite. The stellar kinematics of the GSS in the merger remnant are consistent with observations initially reported by Kalirai et al. (2006) and Gilbert et al. (2007, 2009) and even more so with the results from the wide-field coverage of the inner halo by the DESI survey (D23).
- The R_{proj} versus V_{LOS} diagram for each of the three main substructures in the inner halo of the modelled M31 reveals coherent wedges, chevrons, and stream-like patterns. These features are echoed in the corresponding phase space of resolved stars observed by D23. The two distinct wedge patterns that appear in the phase space of the observed NE and W shelves stars also emerge for modelled stars within the equivalent regions. This two-fold wedge pattern (with a different apocentre for each wedge) does not appear in minor merger models (e.g. Fardal et al. 2007; Kirihara et al. 2017; Milošević et al. 2024). This is justified since a sufficiently massive secondary galaxy is required to drag stellar particles from the pre-merger disc and deposit them in tidal shells.
- The metallicity gradient reported by Conn et al. (2016) and Ogami et al. (2025) along the GSS displaying an increment of the metallicity at 40–50 kpc distance for the GSS is qualitatively reproduced by the model. The model predicts a spread of σ[M/H] in the range of 0.3–0.5 dex which is consistent with the widths of metallicity distributions from spectroscopic measurements but narrower than σ[M/H] values inferred from photometric metallicities (Conn et al. 2016; Ogami et al. 2025).
- The major merger model showcases average [M/H] values for the M31 inner halo substructures that are in agreement

with spectroscopic metallicity observations (Escala et al. 2020; D23). Photometrically estimated (Ogami et al. 2025; Conn et al. 2016; Bernard et al. 2015; Escala et al. 2022; Tanaka et al. 2010) LOS distances, projected distances, and metallicities of stars within the three inner halo substructures are also echoed by the model. In addition, the simulation revealed the effect of the stellar age distribution, which is not accounted for in the current measurements of the metallicity from colour-magnitude diagrams of resolved stellar populations. A known bias of photometric surveys is to select the brightest stars within the GSS and the rest of the substructures, which lie close to the TRGB and thus tend to be younger and more metal rich. The basic assumption of a single age for stellar populations of the satellite galaxy may not be applicable for a massive satellite with an extended star formation history. In particular, the assumption of a single age for the resolved stellar population may break down for the substructures closest to the M31 disc.

Apart from the reproduction of various traits of the M31 stellar halo, the gas-rich major merger model features a reconstructed stellar disc. We obtained the following results for the M31 disc remnant:

- The model qualitatively reproduces the observed metallicity gradient of the M31 disc. The resulting gradients for the two age groups of modelled stars, younger and older than 3.5 Gyr, bracket the metallicity gradients measured for the thin and thicker disc stellar populations.
- The assessment of the radial redistribution of disc stars revealed that the majority of young stars in the model came from the gas particles residing in the outskirts of the progenitors. The initially negative (−0.1 dex/kpc) metallicity gradient led to their lower metal content. Hence, the model provides a viable explanation for the moderately metal-poor, young stars in the outer disc of M31 (see Bernard et al. 2015 and Arnaboldi et al. 2022).
- Future surveys with samples that effectively account for MW contamination without introducing metallicity and ages biases – unlike the selection implemented by D23 to reduce the contamination from MW halo stars – will reveal the younger metal poor outer disc in the W shelf and in the other substructure fields closer to the M31 disc.

Upcoming facilities and instrumentation will focus even more on M31, which is the ideal testbed for resolved stellar population studies of galaxies affected by a recent major merger. The inauguration of facilities with a wide field of view, such as the Roman Space Telescope (see Dey et al. 2023a), and instrumentation, including the Prime Focus Spectrograph (Tamura et al. 2016) on the Subaru Telescope will also allow for resolved stellar population studies in galaxies outside the Local Group. Since mergers are regular events in the evolution of galaxies, the extensive study and juxtaposition against simulated analogues of a major merger remnant such as M31 may be helpful as a guideline for dealing with the upcoming data sets and future modelling efforts.

References

Amorisco, N. C. 2017, MNRAS, 464, 2882

Arnaboldi, M., Bhattacharya, S., Gerhard, O., et al. 2022, A&A, 666, A109

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Bernard, E. J., Ferguson, A. M. N., Barker, M. K., et al. 2012, MNRAS, 420, 2625

Bernard, E. J., Ferguson, A. M. N., Richardson, J. C., et al. 2015, MNRAS, 446, 2789

Bhattacharya, S. 2023, arXiv e-prints [arXiv:2305.03293]

Bhattacharya, S., Arnaboldi, M., Caldwell, N., et al. 2019a, A&A, 631, A56

Bhattacharya, S., Arnaboldi, M., Hartke, J., et al. 2019b, A&A, 624, A132

Bhattacharya, S., Arnaboldi, M., Gerhard, O., et al. 2021, A&A, 647, A130

Bhattacharya, S., Arnaboldi, M., Caldwell, N., et al. 2022, MNRAS, 517, 2343

Bhattacharya, S., Arnaboldi, M., Hammer, F., et al. 2023, MNRAS, 522, 6010

Blaña Díaz, M., Gerhard, O., Wegg, C., et al. 2018, MNRAS, 481, 3210

Brook, C. B., Kawata, D., Gibson, B. K., & Freeman, K. C. 2004, ApJ, 612, 894

Brown, T. M., Smith, E., Ferguson, H. C., et al. 2006, ApJ, 652, 323

Brown, T. M., Smith, E., Ferguson, H. C., et al. 2007, ApJ, 658, L95

Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931

Cohen, R. E., Kalirai, J. S., Gilbert, K. M., et al. 2018, AJ, 156, 230

Conn, A. R., McMonigal, B., Bate, N. F., et al. 2016, MNRAS, 458, 3282

Cox, T. J., Jonsson, P., Primack, J. R., & Somerville, R. S. 2006, MNRAS, 373, 1013

Curti, M., Maiolino, R., Cirasuolo, M., et al. 2020a, MNRAS, 492, 821

Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020b, MNRAS, 491, 944

Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, ApJS, 200, 18

Dalcanton, J. J., Fouesneau, M., Hogg, D. W., et al. 2015, ApJ, 814, 3

Dey, A., Najita, J., Filion, C., et al. 2023a, arXiv e-prints [arXiv:2306.12302]

Dey, A., Najita, J. R., Koposov, S. E., et al. 2023b, ApJ, 944, 1

Dorman, C. E., Guhathakurta, P., Seth, A. C., et al. 2015, ApJ, 803, 24

D'Souza, R., & Bell, E. F. 2021, MNRAS, 504, 5270

Escala, I., Kirby, E. N., Gilbert, K. M., Cunningham, E. C., & Wojno, J. 2019, ApJ, 878, 42

Escala, I., Gilbert, K. M., Kirby, E. N., et al. 2020, ApJ, 889, 177

Escala, I., Gilbert, K. M., Wojno, J., Kirby, E. N., & Guhathakurta, P. 2021, AJ, 162, 45

Escala, I., Gilbert, K. M., Fardal, M., et al. 2022, AJ, 164, 20

Ester, M., Kriegel, H. P., Sander, J., Xu, X., et al. 1996, Second International Conference on Knowledge Discovery and Data Mining (KDD'96), eds. E. Simoudis, J. Han, & U. Fayyad, AAAI Press, 226

Fardal, M. A., Babul, A., Geehan, J. J., & Guhathakurta, P. 2006, MNRAS, 366, 1012

Fardal, M. A., Guhathakurta, P., Babul, A., & McConnachie, A. W. 2007, MNRAS, 380, 15

Fardal, M. A., Babul, A., Guhathakurta, P., Gilbert, K. M., & Dodge, C. 2008, ApJ, 682, L33

Fardal, M. A., Guhathakurta, P., Gilbert, K. M., et al. 2012, MNRAS, 423, 3134

Fardal, M. A., Weinberg, M. D., Babul, A., et al. 2013, MNRAS, 434, 2779

Ferguson, A. M. N., & Mackey, A. D. 2016, in Substructure and Tidal Streams in the Andromeda Galaxy and its Satellites, eds. H. J. Newberg, & J. L. Carlin (Cham: Springer International Publishing), 191

Ferguson, A. M. N., Irwin, M. J., Ibata, R. A., Lewis, G. F., & Tanvir, N. R. 2002, AJ, 124, 1452

Font, A. S., Johnston, K. V., Guhathakurta, P., Majewski, S. R., & Rich, R. M. 2006, AJ, 131, 1436

Gajda, G., Gerhard, O., Blaña, M., et al. 2021, A&A, 647, A131

Garcia, A. M., Torrey, P., Bhagwat, A., et al. 2025, arXiv e-prints [arXiv:2503.03804]

Geehan, J. J., Fardal, M. A., Babul, A., & Guhathakurta, P. 2006, MNRAS, 366, 996

Gilbert, K. M., Fardal, M., Kalirai, J. S., et al. 2007, ApJ, 668, 245

Gilbert, K. M., Guhathakurta, P., Kollipara, P., et al. 2009, ApJ, 705, 1275

Gilbert, K. M., Kirby, E. N., Escala, I., et al. 2019, ApJ, 883, 128

Gilbert, K. M., Wojno, J., Kirby, E. N., et al. 2020, AJ, 160, 41

Gregersen, D., Seth, A. C., Williams, B. F., et al. 2015, AJ, 150, 189

Guhathakurta, P., Rich, R. M., Reitzel, D. B., et al. 2006, AJ, 131, 2497

Hammer, F., Flores, H., Elbaz, D., et al. 2005, A&A, 430, 115

Hammer, F., Flores, H., Puech, M., et al. 2009, A&A, 507, 1313

Hammer, F., Yang, Y. B., Wang, J. L., et al. 2018, MNRAS, 475, 2754

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357

Helmi, A. 2020, ARA&A, 58, 205

Hendel, D., & Johnston, K. V. 2015, MNRAS, 454, 2472

Hopkins, P. F., Hernquist, L., Cox, T. J., Younger, J. D., & Besla, G. 2008, ApJ, 688, 757

Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009, ApJ, 691, 1168

Hubble, E. P. 1929, ApJ, 69, 103

Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90

Ibata, R., Irwin, M., Lewis, G., Ferguson, A. M. N., & Tanvir, N. 2001, Nature, 412, 49

Ibata, R., Chapman, S., Ferguson, A. M. N., et al. 2004, MNRAS, 351, 117

Ibata, R., Chapman, S., Ferguson, A. M. N., et al. 2005, ApJ, 634, 287

Kalirai, J. S., Guhathakurta, P., Gilbert, K. M., et al. 2006, ApJ, 641, 268

Kirihara, T., Miki, Y., & Mori, M. 2014, PASJ, 66, L10

Kirihara, T., Miki, Y., Mori, M., Kawaguchi, T., & Rich, R. M. 2017, MNRAS, 464, 3509

Kobayashi, C., Bhattacharya, S., Arnaboldi, M., & Gerhard, O. 2023, ApJ, 956, L14

Kwitter, K. B., & Henry, R. B. C. 2022, PASP, 134, 022001

Lewis, A. R., Dolphin, A. E., Dalcanton, J. J., et al. 2015, ApJ, 805, 183

Longobardi, A., Arnaboldi, M., Gerhard, O., & Mihos, J. C. 2015, A&A, 579, L3

Maiolino, R., & Mannucci, F. 2019, A&A Rev., 27, 3

Martig, M., Minchev, I., & Flynn, C. 2014, MNRAS, 443, 2452

McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2003, MNRAS, 343, 1335

McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, Nature, 461, 66

McConnachie, A. W., Ibata, R., Martin, N., et al. 2018, ApJ, 868, 55

Merrett, H. R., Kuijken, K., Merrifield, M. R., et al. 2003, MNRAS, 346, L62

Merrett, H. R., Merrifield, M. R., Douglas, N. G., et al. 2006, MNRAS, 369, 120

Merrifield, M. R., & Kuijken, K. 1998, MNRAS, 297, 1292

Milošević, S., Mićić, M., & Lewis, G. F. 2022, MNRAS, 511, 2868

Milošević, S., Mićić, M., & Lewis, G. F. 2024, MNRAS, 527, 4797

Mollá, M., Díaz, Á. I., Cavichia, O., et al. 2019, MNRAS, 482, 3071

Mori, M., & Rich, R. M. 2008, ApJ, 674, L77

Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989

Ogami, I., Tanaka, M., Komiyama, Y., et al. 2025, MNRAS, 536, 530

Okamoto, T., Eke, V. R., Frenk, C. S., & Jenkins, A. 2005, MNRAS, 363, 1299

Pillepich, A., Vogelsberger, M., Deason, A., et al. 2014, MNRAS, 444, 237

Richardson, J. C., Ferguson, A. M. N., Johnson, R. A., et al. 2008, AJ, 135, 1998

Roca-Fàbrega, S., Kim, J.-H., Hausammann, L., et al. 2021, ApJ, 917, 64

Rodrigues, M., Hammer, F., Flores, H., et al. 2008, A&A, 492, 371

Rubin, V. C., Ford, W., & Kent, J. 1970, ApJ, 159, 379

Sadoun, R., Mohayaee, R., & Colin, J. 2014, MNRAS, 442, 160

Saglia, R. P., Opitsch, M., Fabricius, M. H., et al. 2018, A&A, 618, A156

Salaris, M., & Cassisi, S. 2005, Evolution of Stars and Stellar Populations (John Wiley & Sons)

Savino, A., Weisz, D. R., Dolphin, A. E., et al. 2025, ApJ, 979, 205

Stinson, G. S., Bailin, J., Couchman, H., et al. 2010, MNRAS, 408, 812

Tamura, N., Takato, N., Shimono, A., et al. 2016, in Ground-based and Airborne Instrumentation for Astronomy VI, eds. C. J. Evans, L. Simard, & H. Takami, SPIE Conf. Ser., 9908, 99081M

Tanaka, M., Chiba, M., Komiyama, Y., et al. 2010, ApJ, 708, 1168

The pandas development Team 2024, https://doi.org/10.5281/zenodo.10697587

Tissera, P. B., Rosas-Guevara, Y., Bower, R. G., et al. 2019, MNRAS, 482, 2208

Tornatore, L., Borgani, S., Dolag, K., & Matteucci, F. 2007, MNRAS, 382, 1050

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Watkins, L. L., Evans, N. W., & van de Ven, G. 2013, MNRAS, 430, 971

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009, MNRAS, 399, 574

Williams, B. F., Dolphin, A. E., Dalcanton, J. J., et al. 2017, ApJ, 846, 145

Wojno, J. L., Gilbert, K. M., Kirby, E. N., et al. 2023, ApJ, 951, 12

Wyse, R. F. G. 2001, in Galaxy Disks and Disk Galaxies, eds. J. G. Funes, & E. M. Corsini, ASP Conf. Ser., 230, 71

Yin, J., Hou, J. L., Prantzos, N., et al. 2009, A&A, 505, 497

Zinchenko, I. A., Berczik, P., Grebel, E. K., Pilyugin, L. S., & Just, A. 2015, ApJ, 806, 267

Appendix A: Adopted values for the initial metallicity gradient in the two progenitors

To establish the initial metallicity gradient of the two disc progenitors, we relied on constraints derived from observations of galaxies at comparable redshifts. Curti et al. (2020a), utilizing emission line measurements from gravitationally lensed galaxies in the KLEVER survey, determined oxygen abundance gradients within approximately one to two effective radii (R_e). However, we needed to set gradient values that are suitable for the entire radial extent of the discs out to a few tens of kiloparsecs, thus extending beyond the inner $\sim 1-2$ R_e region.

To address this, we considered predictions from chemical evolution models that determine metallicity gradients across the entire radial range of galactic discs. Mollá et al. (2019, see their Figure 5) present the radial distribution of $12 + \log(O/H)$ for MW-like galaxies across redshifts 0 to 4.0. These models exhibit a step-like metallicity profile: a relatively flat gradient in the inner disc regions (up to ~ 10 kpc at redshift $z = 1$), followed by a rapid decline at intermediate radii, and a subsequent flattening in the outer disc (beyond ~ 15 kpc at redshift $z = 1$), where oxygen abundances reach approximately 20% of the solar value. Given our assumption of a linear gradient for the initial discs, we adopt a value of -0.1 dex/kpc. This choice, which is slightly steeper than the near-flat observational gradients reported by Curti et al. (2020a), accounts for the steeper decline predicted by chemical evolution models over the entire radial range covered by the simulated exponential discs in the simulations. Additional support for the adopted gradient value comes from simulated radial metallicity profiles of star-forming galaxies in recent cosmological simulations (see TNG; Garcia et al. 2025 and MUGS; Stinson et al. 2010 cosmological simulations).

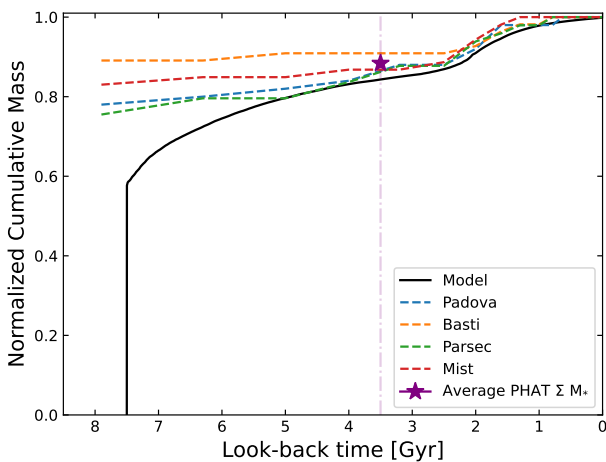## Appendix B: The star formation history of the M31 stellar disc

Fig. B.1. Cumulative stellar mass formed in the north-east quadrant of the disc of M31 from Williams et al. (2017), as probed by four different stellar evolution models (each model is denoted with a dashed coloured line). The purple star indicates the average value for the four models computed 3.5 Gyr ago. The vertical dotted line indicates the separation of the two age groups. The cumulative stellar mass formed during the simulation of the M31 merger event is plotted as a black, solid curve.

Williams et al. (2017) used deep multi-band observations from HST's PHAT survey in the M31 disc to constrain the SFH of its stars. In Figure B.1, we show the SFH using four different stellar evolution models as reproduced in Williams et al. (2017). These results show that $\sim 80\%$ of the stellar mass in the disc is formed at a lookback time of ~ 8 Gyr. At ~ 3.5 Gyr ago, a second star-forming episode started in the observed part of the disc, peaking at $\sim 2-3$ Gyr ago and providing $\sim 20\%$ of the total stellar mass in the M31 disc.

A simplified approach to describing the SFH is to distinguish the M31 disc stars into two age groups. One group includes stars older than 3.5 Gyr, which account for the bulk of the stellar mass in the disc, that is $\sim 80\%$. The second group includes stars born in the subsequent burst of star formation. Stars of this second younger group have ages in the range of 2 to 4 Gyr and younger.

The SFH which results from the H18 simulations deviates somewhat from the M31 SFHs published by Williams et al. (2017) in the PHAT area; the H18 SFH is shown by the black continuous curve in Figure B.1. In the H18 simulations, stars are being formed in the intermediate time range $7.5 - 3.5$ Gyr also. We address such a discrepancy by including all stars older than 3.5 Gyr in the old age group and all stars younger than 3.5 Gyr in the young age group.

We then compare the cumulative mass fraction values for ages <3.5 Gyr and >3.5 Gyr in the simulations with the measured values in the PHAT area. The observed constrained value for the cumulative stellar mass formed for ages older than 3.5 Gyr is $\sim 88\%$ from the average of the four modelled SFHs in Williams et al. (2017). Such value is fully comparable with the cumulative stellar mass faction value generated in the H18 simulation, which amounts to $\sim 84\%$.

Considering these two age families, older and younger than 3.5 Gyr, from the PHAT SFH, we then associate the group of old PNe with the progenitor stars, which are older than 3.5 Gyr. The group of young PNe is then associated with the progenitor stars younger than 3.5 Gyr, which are linked to the recent episode of star formation in the M31 disc.

The age milestone of 3.5 Gyr is in line with the star formation history of the M31's disc as traced by PNe which show two distinct kinematical and chemical populations (Bhattacharya et al. 2022) and with the results from the chemical evolution models in Arnaboldi et al. (2022).

Appendix C: DBSCAN parameters

The parameters that must be specified for the implementation of DBSCAN are the *eps* and *minsamples*. *Eps* is the maximum distance between two samples for one to be considered as being in the neighborhood of the other, and *minsamples* is the number of samples in a neighborhood for a point to be considered as a core point (Ester et al. 1996). The parameters utilised to distinguish the components of each substructure are separately discussed below.

C.1. The Giant Stellar Stream

The GSS in the model consists of regions with diverse densities. The regions that lie in the vicinity of the disc (roughly the area denoted as GSS-component [3] in Figure 6) exhibit a significantly higher density than the rest. DBSCAN clustering does not work efficiently for data sets with varying densities. Therefore, to properly distinguish the constituent components of the GSS, we resort to two consecutive runs of DBSCAN with different values for *eps* and *minsamples*.

For the first run, the values of the parameters were $eps = 3.9$ and *minsamples* = 250. These values are appropriate for the identification of the most dense component, near the disc of the galaxy; GSS-component [3]. We mask this region and resort to the next run of DBSCAN.

For the second run, the values of the parameters were $eps = 4.1$ and *minsamples* = 78. Three more clusters are identified within the premises of the GSS, which, together with the disc component, reveal the GSS's multi-component, complicated intrinsic structure in the model.

C.2. The North-East Shelf

Stars within the NE shelf region selected in the model are almost uniformly spread in density and the DBSCAN parameters used were $eps = 2$ and *minsamples* = 100.

C.3. The Western Shelf

The parameters for the identification of the two components of the (uniformly spread) W shelf are $eps = 2.6$ and *minsamples* = 137.

Appendix D: Line-of-sight modification of the M31 model

H18 chose a specific temporal output during an evolving sequence, which optimises the reproduction of a considerable amount of observations that were previously available, as detailed in Section 2.1. Since the substructures in the inner halo are time-dependent, they may be found closer or further away in the projected distance from their observed counterparts. For instance, H18 reported that some shells can be found at the same radial distance as in M31, but perhaps not simultaneously with other features.

The second limitation is related to the parameters of the final rotation of the simulated analogue at $z = 0$, utilised to place the galaxy in the observed reference frame (PA and inclination of the disc, relative positions of inner halo substructures, etc.). There is some rotation uncertainty in the model since the accuracy of both the PA and the disc inclination is up to a few degrees. Specifically, the third angle that is used to adjust the relative position between the disc and the halo in the simulation is that of the rotation around the disc axis in the plane of the disc. This angle is the least constrained ($\sim \pm 10°$) in reproducing the observational data. Variations in angles impact the V_{LOS} to the MW, as the M31 disc is seen almost edge-on.

The third limitation has to do with the large number of initial parameters (to mention only some of them: the in-fall angle of the secondary galaxy, the value of the first pericentre passage, and the internal light distribution in the progenitors) associated with a major merger simulation. A variation in the initial parameters may provide better agreement with observations; nonetheless, the fine-tuning of the simulation is beyond the scope of this work.

To compare the final temporal output with the observed LOS of the M31 system from our vantage point in the MW, the coordinate system must be fixed. To achieve this, the modelled disc is aligned with the observed inclination and PA of the M31 disc. However, there is a third angle of rotation, the rotation along the disc plane. This angle is termed the kinetic angle. The kinetic angle is accountable for the relative arrangement between the disc and the inner halo substructures. Hence, the chosen kinetic angle is the one that aptly reproduces the observed morphology (i.e. surface brightness) of the GSS and the two shells. However, there is some leeway of approximately $\pm 10°$ in the choice of this kinetic angle while maintaining the GSS (and other substructures) surface brightness similar to the observed.

To illustrate the extent of the impact on the phase space diagram introduced from choosing a different kinetic angle, Figure D.1 depicts the R_{proj} versus V_{LOS} diagram of stars in the GSS for a slightly different kinetic angle ($+7°$), which still reproduces the observed morphology of the inner halo features. The density of stars in the identified clusters is significantly altered compared to that observed for the default kinetic angle (compare with Figure 7). This is explicitly noticeable in the reduced density of GSS-component [3], which impacts the identification of coherent features in the phase space.

Therefore, the adjustment of the LOS alignment of the simulated galaxy at $z = 0$ may provide a viable mechanism to account for discrepancies between the observed and modelled phase space features in M31's substructures.
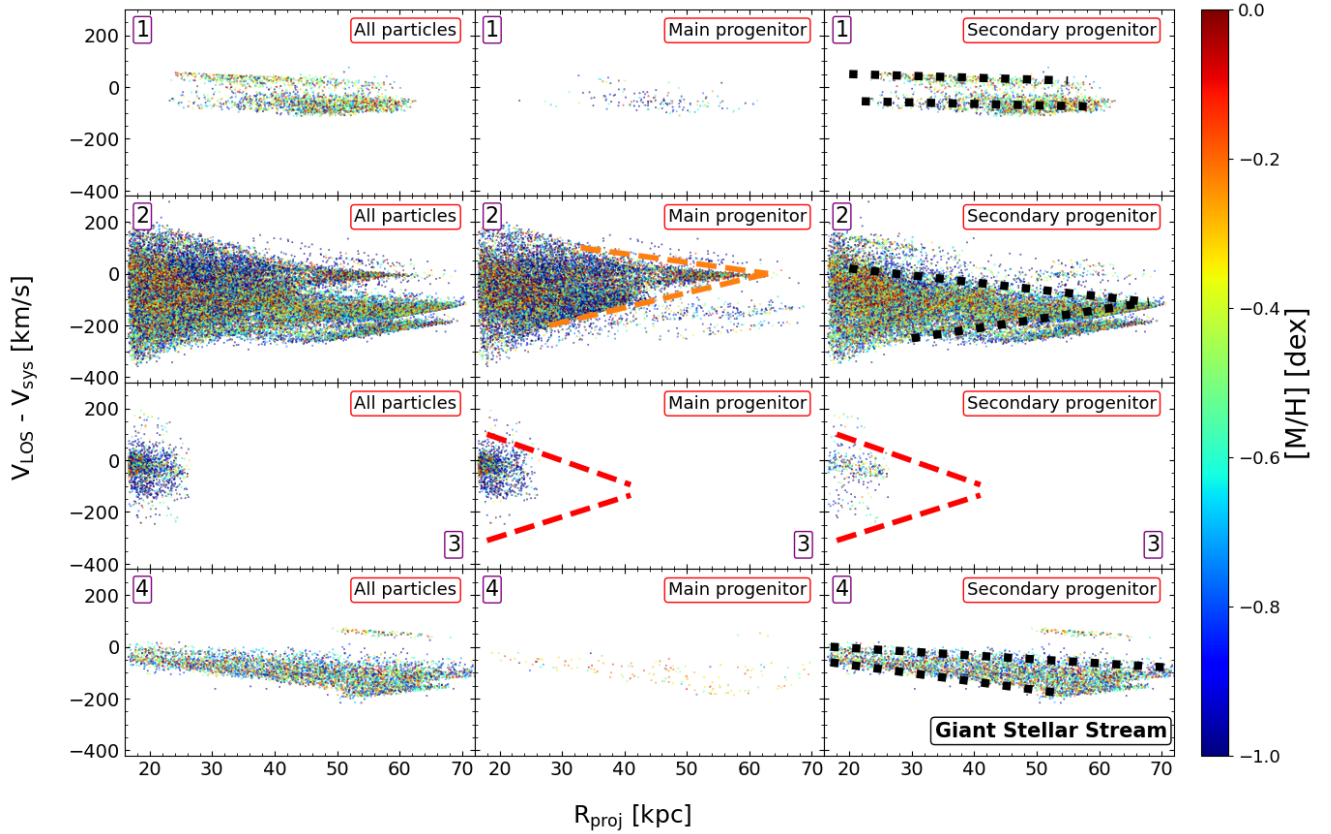
Fig. D.1. Same as Figure 7, but for a different kinetic angle. Delineated ridges are the ones previously identified for the GSS with the default kinetic angle, to explicitly show the differences that emerge from choosing a different kinetic angle.

Appendix E: Comparing photometric metallicity of substructures with younger model population

There is a degeneracy inherent in determining the age and metallicity of a stellar population from photometric observations. A single value of the age of the stellar population has been assumed to obtain the corresponding metallicity and distance from the MW for the M31 inner halo substructures from their CMDs (Conn et al. 2016; Ogami et al. 2025). As shown in Figure 14, the GSS spans a distance of ~160 kpc along the LOS. Hence, determining a unique distance modulus for the GSS may bias its determined metallicity. Furthermore, a photometric survey from the entirety of stars stretching along ~160 kpc LOS distance will preferentially sample the brightest resolved giants, therefore, those near the tip of the RGB. These stars, however, are in principle the younger RGBs. Therefore, photometric studies to determine stellar metallicity in the GSS may entail a possible bias towards younger RBG stars. To examine such a potential bias, we replicate Figure 14 but selecting only model stars younger than 3.5 Gyr (Figure E.1). We find that selecting only young (<3.5 Gyr old) stars in the model results in a better agreement with photometric metallicity observations in the GSS. For the NE and W shelves, selecting younger stars that are <3.5 Gyr old in the model leads to average metallicity values that are slightly more metal-rich than the metallicity values determined from photometric observations for these substructures, obtained assuming an age of 12 − 13 Gyr (see Figures E.2 and E.3).
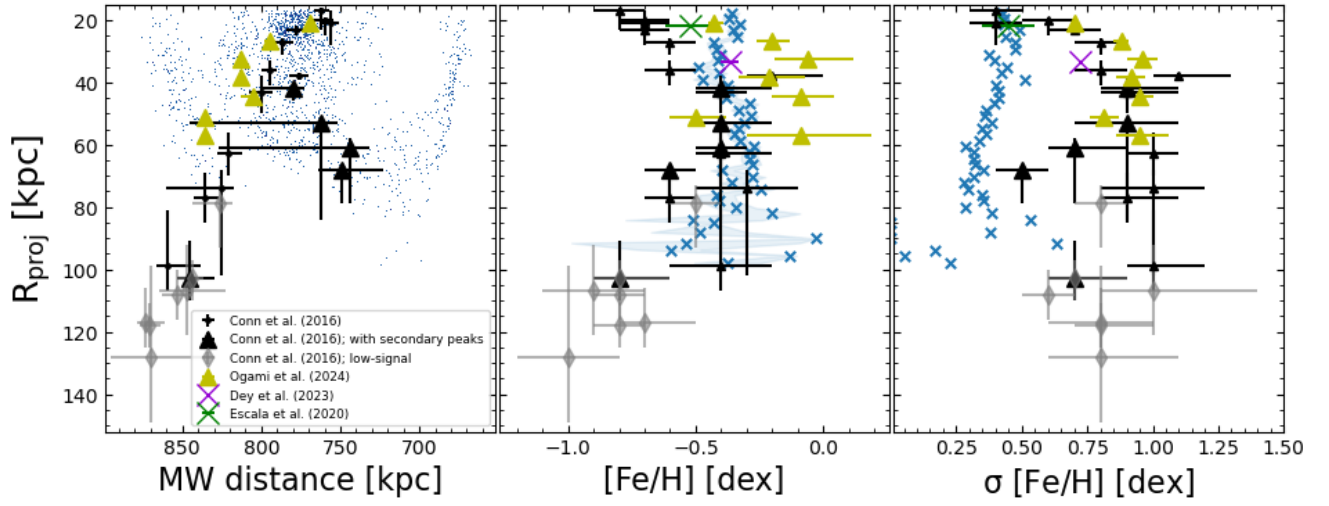
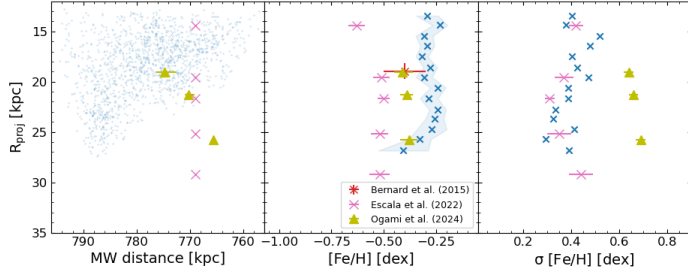**Fig. E.1.** Same as Figure 14 but for modelled stars younger than 3.5 Gyr.

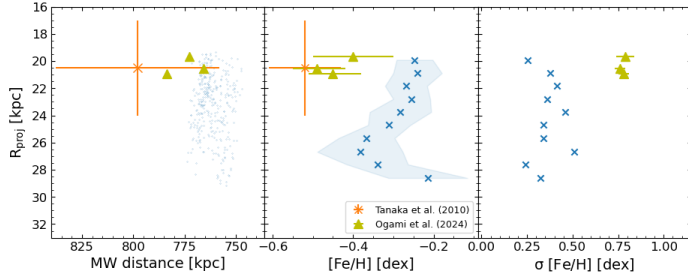**Fig. E.2.** Same as Figure 15 but for modelled stars younger than 3.5 Gyr.

Fig. E.3. Same as Figure 16 but for modelled stars younger than 3.5 Gyr.